GEOMETRIC APPROACH TO HAMILTONIAN DYNAMICS AND STATISTICAL MECHANICS

Lapo CASETTI[a], Marco PETTINI[b], E.G.D. COHEN[c]

[a] *Istituto Nazionale per la Fisica della Materia (INFM), Unità di Ricerca del Politecnico di Torino, Dipartimento di Fisica, Politecnico di Torino, Corso Duca degli Abruzzi 24, I-10129 Torino, Italy*
[b] *Osservatorio Astrofisico di Arcetri, Largo Enrico Fermi 5, I-50125 Firenze, Italy, and Istituto Nazionale per la Fisica della Materia (INFM), Unità di Ricerca di Firenze, Largo Enrico Fermi 2, I-50125 Firenze, Italy*
[c] *The Rockefeller University, 1230 York Avenue, New York, 10021-6399, USA*

ELSEVIER

AMSTERDAM – LAUSANNE – NEW YORK – OXFORD – SHANNON – TOKYO

ELSEVIER

Physics Reports 337 (2000) 237–341

PHYSICS REPORTS

www.elsevier.com/locate/physrep

Geometric approach to Hamiltonian dynamics and statistical mechanics

Lapo Casetti[a,*], Marco Pettini[1b], E.G.D. Cohen[c]

[a]*Istituto Nazionale per la Fisica della Materia (INFM), Unità di Ricerca del Politecnico di Torino, Dipartimento di Fisica, Politecnico di Torino, Corso Duca degli Abruzzi 24, I-10129 Torino, Italy*

[b]*Osservatorio Astrofisico di Arcetri, Largo Enrico Fermi 5, I-50125 Firenze, Italy, and Istituto Nazionale per la Fisica della Materia (INFM), Unità di Ricerca di Firenze, Largo Enrico Fermi 2, I-50125 Firenze, Italy*

[c]*The Rockefeller University, 1230 York Avenue, New York, 10021-6399, USA*

Received April 2000; editor: I. Procaccia

Contents

* Corresponding author. Present address: INFM, UdR Firenze, Dipartimento di Fisica, Università di Firenze, Largo Enrico Fermi 2, I-50125 Firenze, Italy. Tel.: + 39-055-2307850; fax: + 39-055-229330.

E-mail addresses: casetti@fi.infn.it (L. Casetti), pettini@arcetri.astro.it (M. Pettini), egdc@rockefeller.edu (E.G.D. Cohen).

[1] Also at INFN, Sezione di Firenze, Largo Enrico Fermi 2, I-50125 Firenze, Italy.

PII: S0370-1573(00)00069-7

Abstract

This paper is a review of results which have been recently obtained by applying mathematical concepts drawn, in particular, from differential geometry and topology, to the physics of Hamiltonian dynamical systems with many degrees of freedom of interest for statistical mechanics. The first part of the paper concerns the applications of methods used in classical differential geometry to study the chaotic dynamics of Hamiltonian systems. Starting from the identity between the trajectories of a dynamical system and the geodesics in its configuration space, when equipped with a suitable metric, a geometric theory of chaotic dynamics can be developed, which sheds new light on the origin of chaos in Hamiltonian systems. In fact, it appears that chaos can be induced not only by negative curvatures, as was originally surmised, but also by positive curvatures, provided the curvatures are fluctuating along the geodesics. In the case of a system with a large number of degrees of freedom it is possible to approximate the chaotic instability behaviour of the dynamics by means of a geometric model independent of the dynamics, which allows then an analytical estimate of the largest Lyapunov exponent in terms of the averages and fluctuations of the curvature of the configuration space of the system. In the second part of the paper the phenomenon of phase transitions is addressed and it is here that topology comes into play. In fact, when a system undergoes a phase transition, the fluctuations of the configuration-space curvature, when plotted as a function of either the temperature or the energy of the system, exhibit a singular behaviour at the phase transition point, which can be qualitatively reproduced using geometric models. In these models the origin of the singular behaviour of the curvature fluctuations appears to be caused by a topological transition in configuration space, which corresponds to the phase transition of the physical system. This leads us to put forward a *topological hypothesis* (TH). The content of the TH is that phase transitions would be related at a deeper level to a change in the topology of the configuration space of the system. We will illustrate this on a simple model, the mean-field XY model, where the TH can be checked directly and analytically. Since this model is of a rather special nature, namely a mean-field model with infinitely ranged interactions, we discuss other more realistic (non-mean-field-like) models, which cannot be solved analytically, but which do supply direct supporting evidence for the TH via numerical simulations.

PACS: 02.40. − k; 05.20. − y; 05.45. + b; 05.70.Fh

To the memory of Lando Caiani

1. Introduction

This paper deals with the application of concepts drawn from mathematics, in particular from differential geometry and topology, to problems in statistical physics. The mathematical tools involved come from Riemannian geometry and from Morse theory, respectively. As to the physics, the applications of these concepts will be brought to bear on dynamical systems with many degrees of freedom, including eventually the thermodynamic limit.

In order to contain this report to a reasonable size and yet make it accessible to as wide a readership as possible, and since it makes use of concepts which might not be known to everyone, we chose the following format.

The first part of the main text is aimed at a reader who is familiar with the basics of Riemannian geometry, for example at the level of a course in general relativity. As to the second part, the knowledge of Morse theory at an elementary level is assumed. However, for those who are not familiar with these branches of mathematics, we have provided in extended appendices the main points which are needed to follow the exposition. Similarly, we assume that the reader is familiar with the basics of dynamical systems theory, but again we summarized in an appendix the main concepts. In all cases references to the literature are made for the details. Both the main text and the appendices are written as a compromise between mathematical rigour and a physicist's accessibility; in case of conflict, we always favoured the latter.

This way we hope that a reader familiar with the basic mathematical tools will be able to read the report straightforwardly. We have made a special effort to emphasize logical structure and physical content, and we hope that the report will provide a clear survey of what has been achieved applying geometrical methods to dynamical systems and statistical mechanics. At the same time we would like this paper to allow the reader to familiarize herself or himself with this new field and to stimulate new developments and contributions to the many points which are still open.

Throughout the paper we will consider classical Hamiltonian dynamical systems with N degrees of freedom, confined in a finite volume,[2] whose Hamiltonian is of the form

$$\mathscr{H} = \frac{1}{2}\sum_{i=1}^{N} p_i^2 + V(q_1, \ldots, q_N), \tag{1}$$

where the q's and the p's are, respectively, the coordinates and the conjugate momenta of the system. Our emphasis is on systems with a large number of degrees of freedom. The dynamics of system (1) is defined in the $2N$-dimensional phase space spanned by the q's and the p's. Our aim is to relate the dynamical and the statistical properties of system (1) with the geometrical and topological properties of the phase space where the dynamical trajectories of the system live. It turns out that as long as we consider Hamiltonians of the form (1) we can restrict ourselves to the study of the geometry and the topology of the N-dimensional configuration space[3] without loosing

[2] We will mainly apply our results to systems defined on a lattice, so that we will not explicitly refer to the volume of the system.

[3] Actually, we will also consider an enlarged configuration space with two extra dimensions.

information. In fact as we shall see in Section 3 the dynamical trajectories can be seen as geodesics of the configuration space, provided the latter has been endowed with a suitable metric. As to the topology, later on we shall see that also all the relevant information on the topology of phase space, from the point of view of Morse theory, is encoded in the potential energy function $V(q_1, \ldots, q_N)$, so that also the topological investigation can be restricted to the configuration space.

This is similar to what happens in the classical statistical mechanics of Hamiltonian systems of the form (1), where the momenta can be integrated out and the statistical measure can be defined on the configuration space alone. We remark that this is true for both the microcanonical and the canonical ensemble.

The structure of the paper is the following: after a short historical introduction (Section 2), the main body of the paper is organized in two parts, which, though being tightly related, are also to a great extent independent from each other, so they could be also read separately without encountering too many difficulties.

The first part (Sections 3 and 4) concerns the applications of tools belonging to classical differential geometry to study the chaotic dynamics of Hamiltonian systems of the form (1), in particular for those with a large number of degrees of freedom. Starting from the identity between the trajectories of the dynamical system (1) and the geodesics of the configuration space equipped with the Jacobi or the Eisenhart metric, we develop a geometric theory of chaotic dynamics which sheds a new light on the origin of chaos in Hamiltonian systems. In fact, it turns out that chaos can be induced not only by negative curvatures, but also by positive curvatures, provided the curvatures are fluctuating along the geodesics. In the case of a large number of degrees of freedom it is possible to describe the instability of the dynamics by means of a geometric model independent of the dynamics, which provides an analytical estimate of the largest Lyapunov exponent in terms of the averages and fluctuations of the curvature of the configuration space. The basic difficulty in doing this is the extremely complicated and unknown nature of the configuration space geometry of many particle systems. To obtain concrete results, we had to introduce a number of simplifying assumptions on physical grounds, which seem, however, to capture some of the essentials of the configuration space geometry, judging from the good agreement obtained with computer simulations for the largest Lyapunov exponent. Conversely, this gives us, in principle, some insight in the geometry of configuration space relevant for the computation of observable properties of many particle systems. We clarified the exposition with the application of the general concepts to two special examples; however, since some of the calculations are rather lengthy, we did not provide all the details, referring the reader for those to the appropriate literature.

While in the first part we deal with the application of differential geometry to dynamical systems with a large number of degrees of freedom, we do not touch upon one of the most spectacular properties of large systems, namely that in the thermodynamic limit $N \to \infty$ they may show sharp phase transitions. In the second part (Sections 5 and 6) we address this point, and it is here that topology comes into play. In fact, when a system undergoes a phase transition, the fluctuations of the configuration-space curvature, when plotted as a function of either the temperature or the energy, have a singular behaviour at the transition point which can be qualitatively reproduced using a geometric model. In such a model the origin of the singular behaviour of the curvature fluctuations resides in a topological change. This leads us to put forward a *topological hypothesis* (TH). The content of the TH is that phase transitions (at least, continuous phase transitions) would at a deeper level be related to a particular change in the topology of the configuration space of the

system. We will illustrate this on a simple model, the mean-field XY model, where the TH can be checked directly and analytically. Since this model is of a rather special nature, namely a mean-field model with infinitely ranged interactions, we discuss other more realistic (non-mean-field-like) models, which cannot be solved analytically, but which do supply indirect supporting evidence for the TH via numerical simulations.

We emphasize that the purpose of the work discussed in the second part of this report is not to extend the existing theory of phase transitions for systems described by a classical Hamiltonian. Rather, we try to extend the foundation for the occurrence of phase transitions to transitions in the topology of the configuration space of the system undergoing a phase transition. This way we hope to make a new connection between a branch of pure mathematics (topology) and a branch of statistical mechanics (phase transitions). Such a connection appears to lead to a new approach to phase transitions.

A brief historical summary, to place the content of the present review in its historical context, is given in the next section.

2. Historical remarks

Without attempting to be exhaustive, a few historical comments might be helpful to place the recent contributions about the geometrical approach to dynamics and statistical physics which are reviewed in the present article, in a more general context. This makes the present section an exception to the self-containedness of this review paper, because we mention here concepts which are not necessary to understand the topics treated in the rest of the paper.

The idea of looking at the collection of solutions of Newton's equations of motion from a geometric point of view dates back to Poincaré and to the development of the qualitative theory of differential equations. Tackling the famous problem of the integrability of the three-body problem, Poincaré also discovered that generic classical Hamiltonian systems, in spite of their deterministic nature, lack predictability, i.e. were unstable, because of their extreme sensitivity to the initial conditions. Such an instability of classical dynamics originates in homoclinic intersections, which Poincaré described in his *Méthodes nouvelles de la mécanique céleste* [1] without "even attempting to draw" them (see Appendix C). However his geometric treatment of dynamics, later developed by Cartan among others, involves submanifolds of phase space using what is now called symplectic geometry [2]. Although of undeniable elegance, symplectic geometry is not very helpful to advance our knowledge about those regions in phase space where the dynamics is unstable, knowledge which is relevant for statistical mechanics.

The name of Poincaré, together with that of Fermi, is also associated with an important theorem about the non-existence of analytic integrals of motion, besides energy, for generic non-linear Hamiltonian systems describing at least three interacting bodies [1,3]; this is the origin of the accessibility of the whole constant energy hypersurface of phase space which is determined by the initial conditions. The Poincaré–Fermi theorem has been generally considered by the physics community as sufficient to legitimate classical statistical mechanics from a dynamical viewpoint.[4]

[4] Kol'mogorov–Arnol'd–Moser (KAM) theory [4] might seem capable of explaining how, in spite of this "no-go" theorem due to Poincaré and Fermi, ergodicity could fail. However, the exceedingly tiny – and fastly vanishing with the number N of degrees of freedom – non-integrable perturbations that are required to keep a positive measure of regular regions in phase space, do, in general, not have any physically appreciable effect even when only a few bodies are considered.

One had to wait until the 1940s when a qualitatively new attempt emerged to make use of geometric concepts in the investigation of newtonian dynamics and its connection with statistical mechanics. Though very different from the clear mathematical expositions of Poincaré, and much more in the spirit of a physicist, it was Krylov [5] in his doctoral dissertation, who showed for the first time the existence of a close relationship between dynamical instability (seen as the exponential amplification of small deviations in the initial conditions of a collection of colliding objects representing idealized atoms in a gas) and phase space mixing. Phase mixing is a stronger property than ergodicity and is far more relevant to physics than ergodicity. In fact, while ergodicity assures the equality of time and phase space averages of physical quantities, phase mixing addresses the rate of approach in time to ensemble averages. Phase mixing entails thus the convergence of time averages to ensemble averages in a finite time. In modern terms, Krylov realized the necessity of chaotic dynamics to obtain phase mixing and to make the connections between dynamics and statistical mechanics stronger. Moreover Krylov, in view of what we are going to discuss in the next sections, also has the great historical merit of having attempted for the first time to bridge the dynamical foundations of statistical mechanics with a widely developed and powerful field of mathematics: Riemannian differential geometry. Krylov knew certain mathematical results concerning the properties of geodesic flows on compact negatively curved manifolds by Hadamard et al. [6], and he realized their potential interest to physics, once Newtonian dynamics is rephrased in terms of Riemannian geometric language. Such a possibility was well known since the beginning of the century, mainly due to the work of Levi–Civita, in particular that the principle of stationary action entails the identity of a classical mechanical flow with a geodesic flow in a configuration space, endowed with a suitable metric. Krylov's efforts concentrated on the analysis of the properties of physical systems which move in negatively curved regions in configuration space. For example, he discussed how the presence of an inflection point in the Lennard–Jones potential could influence the dynamics of a dilute gas (through the appearance of regions of negative scalar curvature in configuration space) and lead to a strong instability of the dynamics. These attempts have been very influential on the development of the so-called abstract ergodic theory [7], where Anosov flows [8] (e.g., geodesic flows on compact manifolds with negative curvature) play a prominent role. Ergodicity and mixing of these flows have been thoroughly investigated. To give an example, Sinai proved ergodicity and mixing for two hard spheres by just showing that such a system is equivalent to a geodesic flow on a negatively curved compact manifold [9].

From time to time, Krylov's intuitions have been worked out further by several physicists, for whom we refer to Refs. [10–20]. They invariably discovered, much to their surprise, that geodesic flows associated with physical Hamiltonians do not live on negatively curved manifolds, despite their chaoticity even if the latter is well developed; only very few exceptions to this are known, in fact, the two low-dimensional models discussed in Refs. [12,15], where chaos is actually associated with negative curvature. Worse, for certain models the regions of negative curvature of the mechanical manifolds apparently shrink by increasing the number N of degrees of freedom.

This somewhat biased search for negative curvatures has been the main obstacle to an effective use of the geometric framework originated by Krylov to explain the source of chaos in Hamiltonian systems. On the other hand, it is true that the Jacobi equation, which describes the (in)stability of a geodesic flow, is in practice only tractable on negatively curved manifolds, formidable mathematical difficulties are encountered in treating the (in)stability of geodesic flows on manifolds of

non-constant and not everywhere negative curvature. Moreover, for this kind of problems, intuition can hardly help. However, the advent of computers has been here of invaluable help. In this connection, it may not be out of place to quote here some of the sentences which S. Ulam remembered from the far-looking conversations he had with Fermi and von Neumann [21]:

> After the war, during one of his frequent summer visits to Los Alamos, Fermi became interested in the development and potentialities of the electronic computing machines. He held many discussions with me on the kind of future problems which could be studied through the use of such machines. We decided to try a selection of problems for heuristic work where in the absence of closed analytic solutions experimental work on a computing machine would perhaps contribute to the understanding of properties of solutions.
> (...) Fermi expressed often the belief that future fundamental theories in physics may involve nonlinear operators and equations, and that it would be useful to attempt practice in the mathematics needed for the understanding of nonlinear systems (...).

As a matter of fact, only during the last few years an interplay between analytic methods and numerical simulations has made it possible to overcome the mentioned difficulties, proving the effectiveness of the Riemannian geometric approach to dynamical systems of interest to statistical mechanics, field theory, and condensed matter physics [22–39]. As we shall see in the following, this has extended the domain of application of geometric techniques, and has also introduced a new point of view about the origin of chaos in Hamiltonian systems, as well as new methods to describe and understand it.

The use of Finsler manifolds (generalizations of Riemannian manifolds that allow the geometrization of velocity-dependent potentials as well) has also been proposed in Refs. [40,41]. An analysis of dynamics based on the geometry of trajectories, rather than of the manifolds on which they move, has been proposed in Ref. [42].

For what concerns the use of geometric and topologic concepts in statistical mechanics, we must distinguish between macroscopic and microscopic phase spaces. A macroscopic phase space is a low-dimensional space spanned by macroscopic variables, like temperature, pressure, volume, chemical potential, etc., in other words, it is in general a parameter space. In the 1970s some applications to the study of phase transitions of the theory of singularities of differentiable maps (popularly known as Catastrophe theory), which includes Morse theory, were proposed. These followed Thom's remark that the simplest example of the classical critical point as it appears in the van der Waals equation corresponds to the Riemann–Hugoniot catastrophe [43,44].

An elegant and deep formulation of phase transitions related to structural instability and using one of the most beautiful theorems in differential topology, the Atiyah–Singer index theorem, was proposed by Rasetti in Ref. [45].

Other very recent proposals of geometric and topologic methods in macroscopic phase spaces have been put forward in Refs. [46,47].

In recent papers some elements of the geometry of constant energy hypersurfaces Σ_E of phase space have been used for the microcanonical definitions of the temperature and the specific heat [48,49], in Ref. [50] a topological property of the Σ_E has been related to the mean curvature of the Σ_E from which a relationship between thermodynamics and topology emerged.

The description of phase transitions through geometric and topologic changes in the microscopic phase space has never been considered until very recently. It appeared as a natural consequence of the above-mentioned new developments in the Riemannian theory of Hamiltonian chaos. These newly proposed ideas, as well as the conceptual path that led from the geometry of dynamics to topology and phase transitions, are reviewed in this paper.

3. Riemannian geometry and Hamiltonian dynamics

A Hamiltonian system whose kinetic energy is a quadratic form in the velocities is referred to as a *natural* Hamiltonian system. Every Newtonian system, that is a system of particles interacting through forces derived from a potential, i.e. of the form (1), belongs to this class. The trajectories of a natural system can be seen as geodesics of a suitable Riemannian manifold. This classical result is based on a variational formulation of dynamics. In fact, Hamilton's principle states that the natural motions of a Hamiltonian system are the extrema of the functional (Hamiltonian action $\mathscr{S}$)

$$\mathscr{S} = \int \mathscr{L}\, \mathrm{d}t\,, \tag{2}$$

where $\mathscr{L}$ is the Lagrangian function of the system, and the geodesics of a Riemannian manifold are the extrema of the length functional

$$\ell = \int \mathrm{d}s\,, \tag{3}$$

where s is the arc-length parameter. Once a connection between length and action is established, by means of a suitable choice of the metric, it will be possible to identify the geodesics with the physical trajectories.

3.1. Geometric formulation of the dynamics

The Riemannian formulation of classical dynamics is far from unique, even if we restrict ourselves to the case of natural systems. There are many possible choices for the ambient space and its metric. The most commonly known choice – dating back to the 19th century – is the so-called Jacobi metric on the configuration space of the system. Actually, this was the geometric framework of Krylov's work. Among other possibilities, we will also consider a metric originally introduced by Eisenhart on an enlarged configuration space–time. The choice of the metric to be used will be dictated mainly by convenience.

These choices certainly do not contain all the possibilities of geometrizing conservative dynamics. For instance, with regards to systems whose kinetic energy is not quadratic in the velocities – the classical example is a particle subject to conservative as well as velocity-dependent forces, such as the Lorentz force – it is impossible to give a Riemannian geometrization, but becomes possible in the more general framework of a Finsler geometry [51]. However, we will not consider this here, and restrict ourselves to natural Hamiltonian systems.

For a summary of the notation and the concepts of differential geometry that will be used in the following we refer the reader to Appendix A. The summation convention over repeated indices will be always used throughout the paper, if not explicity stated otherwise.

3.1.1. The Jacobi metric

Let us consider an autonomous dynamical system, i.e., a system with interactions which do not explicitly depend on time, whose Lagrangian can be written as

$$\mathscr{L} = T - V = \tfrac{1}{2}a_{ij}\dot{q}^i\dot{q}^j - V(q), \tag{4}$$

where the dot stands for a derivative with respect to the parameter on which the q's depend,[5] and q is a shorthand notation for all the coordinates $q_1, \ldots, q_N$. Both these conventions will be used throughout the paper, when there is no possibility of confusion.

The Hamiltonian $\mathscr{H} = T + V$ is an integral of motion, whose value, the energy E, is a conserved quantity. Hence, Hamilton's priciple can be cast in Maupertuis' form [2]: the natural motions of the system are the stationary paths in the configuration space M for the functional

$$\mathscr{A} = \int_{\gamma(t)} p_i \, \mathrm{d}q^i = \int_{\gamma(t)} \frac{\partial \mathscr{L}}{\partial \dot{q}^i} \, \dot{q}^i \, \mathrm{d}t \tag{5}$$

among all the isoenergetic curves, i.e. the curves $\gamma(t)$ connecting the initial and final points parametrized so that the Hamiltonian $\mathscr{H}(p, q)$ is a constant equal to the energy E. The fact that the curves must be isoenergetic with energy E implies that the accessible part of the configuration space is not the whole M, but only the subspace $M_E \subset M$ defined by

$$M_E = \{q \in M\colon V(q) \leq E\}. \tag{6}$$

In fact, a curve γ' which lies outside M_E will never be parametrizable in such a way that the energy is E, because γ' will then pass through points where $V > E$ and the kinetic energy is positive.[6]

The kinetic energy T is a homogeneous function of degree two in the velocities, hence Euler's theorem implies that

$$2T = \dot{q}^i \frac{\partial \mathscr{L}}{\partial \dot{q}^i} \tag{7}$$

and Maupertuis' principle reads as

$$\delta \mathscr{A} = \delta \int 2T \, \mathrm{d}t = 0. \tag{8}$$

The configuration space M of a dynamical system with N degrees of freedom has a differentiable manifold structure, and the Lagrangian coordinates $(q_1, \ldots, q_N)$ can be regarded as local coordinates on M. The latter becomes a Riemannian manifold once a proper metric is defined. For the

[5] Such a parameter is the time t here, but could also be the arc-length s in the following.

[6] The accessible configuration space M_E can then be seen as the union of all the "sub-configuration spaces" $\{q \in M\colon V(q) = E - T\}$ that one gets for all the possible values of T, $0 \leq T \leq E$.

sake of simplicity, let us consider systems of the form (1), so that the kinetic energy matrix is diagonal and the masses are all equal to one, i.e., $a_{ij} = \delta_{ij}$. If we write

$$g_{ij} = 2[E - V(q)]\,\delta_{ij}\,, \tag{9}$$

then Eq. (8) becomes

$$0 = \delta \int 2T\,\mathrm{d}t = \delta \int (g_{ij}\dot{q}^i\dot{q}^j)^{1/2}\,\mathrm{d}t = \delta \int \mathrm{d}s\,, \tag{10}$$

so that the natural motions are the geodesics of M provided $\mathrm{d}s$ is the arc-length element, i.e., the metric on M is given by the tensor whose components are just the g_{ij} defined in Eq. (9). This metric is referred to as the *Jacobi metric*, and its arc-length element is

$$\mathrm{d}s^2 \equiv g_{ij}\,\mathrm{d}q^i\,\mathrm{d}q^j = 2[E - V(q)]\frac{\mathrm{d}q^i}{\mathrm{d}t}\frac{\mathrm{d}q_i}{\mathrm{d}t}\,\mathrm{d}t^2 = 4[E - V(q)]^2\,\mathrm{d}t^2\,. \tag{11}$$

The Jacobi metric vanishes (i.e., is singular[7]) on the boundary ∂M_E of the manifold M_E (∂M_E is often referred to as the classical turning manifold, and is the locus where the trajectories are reflected).

The geodesic equations written in the local coordinate frame $(q^1, \ldots, q^N)$ are (see Eq. (A.34))

$$\frac{\mathrm{D}\dot{\gamma}}{\mathrm{d}s} \equiv \frac{\mathrm{d}^2q^i}{\mathrm{d}s^2} + \Gamma^i_{jk}\frac{\mathrm{d}q^j}{\mathrm{d}s}\frac{\mathrm{d}q^k}{\mathrm{d}s} = 0\,, \tag{12}$$

where $D/\mathrm{d}s$ is the covariant derivative along the curve $\gamma(s)$ (see Eqs. (A.27) and (A.32)), $\dot{\gamma} = \mathrm{d}q/\mathrm{d}s$ is the velocity vector of the geodesic and the Γ are the Christoffel symbols. Using the definition of the Christoffel symbols (see Eqs. (A.31)) and (9) it is straightforward to show that the Eqs. (12) become

$$\frac{\mathrm{d}^2q^i}{\mathrm{d}s^2} + \frac{1}{2(E-V)}\left[2\frac{\partial(E-V)}{\partial q_j}\frac{\mathrm{d}q^j}{\mathrm{d}s}\frac{\mathrm{d}q^i}{\mathrm{d}s} - g^{ij}\frac{\partial(E-V)}{\partial q_j}g_{km}\frac{\mathrm{d}q^k}{\mathrm{d}s}\frac{\mathrm{d}q^m}{\mathrm{d}s}\right] = 0\,, \tag{13}$$

whence, using Eq. (11), Newton's equations are recovered,

$$\frac{\mathrm{d}^2q^i}{\mathrm{d}t^2} = -\frac{\partial V}{\partial q_i}\,. \tag{14}$$

Note that the Jacobi metric is obtained by a conformal change of the kinetic energy metric a_{ij} – see Eq. (9) and Appendix A.3. In fact, the general result for the Riemannian geometrization of natural Hamiltonian dynamics is the following:

Theorem. *Given a dynamical system on a Riemannian manifold* (M, a), *i.e., a dynamical system whose Lagrangian is*

$$\mathscr{L} = \tfrac{1}{2}a_{ij}\dot{q}^i\dot{q}^j - V(q)\,,$$

[7] This singularity might affect in a non-trivial fashion the stability of the trajectories (see e.g. Ref. [52]). However, reflections of the trajectory at the boundary become very unlikely in the case of systems with a very large number of degrees of freedom.

then it is always possible to find a conformal transformation of the metric,

$$g_{ij} = \mathrm{e}^{\varphi(q)} a_{ij}$$

such that the geodesics of (M, g) *are the trajectories of the original dynamical system*; *this transformation is defined by*

$$\varphi(q) = \log[E - V(q)] .$$

The proof proceeds as above, using Eqs. (9)–(14) and simply replacing all the δ_{ij} matrices with the kinetic energy matrix a_{ij}; for details, see e.g. Ref. [11].

3.1.2. The Eisenhart metric

We could try to consider the configuration spacetime $M \times \boldsymbol{R}$ as an alternative ambient space for the geometrization of dynamics, with local coordinates $(q^0 = t, q^1, \ldots, q^N)$, and to define a metric starting from Hamilton's principle $\delta \int \mathscr{L}\, \mathrm{d}t = 0$. We could try to define a metric tensor by multiplying the Lagrangian (4) by $2(\mathrm{d}q^0)^2$

$$\mathrm{d}s^2 = 2\mathscr{L}\,(\mathrm{d}q^0)^2 = (g_{\mathscr{L}})_{\mu\nu}\, \mathrm{d}q^\mu\, \mathrm{d}q^\nu = a_{ij}\, \mathrm{d}q^i\, \mathrm{d}q^j - 2V(q)\,(\mathrm{d}q^0)^2 , \tag{15}$$

where μ and ν run from 0 to N and i and j run from 1 to N; however, one can easily verify that the geodesics of the manifold $(M \times \boldsymbol{R}, g_{\mathscr{L}})$ are then not the natural motions of the systems, since the Lagrangian is not an integral of the motion.

However, we can consider an ambient space with an extra dimension, $M \times \boldsymbol{R}^2$, with local coordinates $(q^0, q^1, \ldots, q^i, \ldots, q^N, q^{N+1})$. This space can be endowed with a non-degenerate pseudo-Riemannian metric (see Appendix A.1.2), first introduced by Eisenhart [53], whose arc-length is

$$\mathrm{d}s^2 = (g_E)_{\mu\nu}\, \mathrm{d}q^\mu\, \mathrm{d}q^\nu = a_{ij}\, \mathrm{d}q^i\, \mathrm{d}q^j - 2V(q)(\mathrm{d}q^0)^2 + 2\,\mathrm{d}q^0\, \mathrm{d}q^{N+1} , \tag{16}$$

where μ and ν run from 0 to $N+1$ and i and j run from 1 to N, and which, from now on, will be referred to as the *Eisenhart metric*, and whose metric tensor will be denoted as g_E. The relation between the geodesics of this manifold and the natural motions of the dynamical system is contained in the following [54]:

Theorem (Eisenhart)**.** *The natural motions of a Hamiltonian dynamical system are obtained as the canonical projection of the geodesics of* $(M \times \boldsymbol{R}^2, g_E)$ *on the configuration space-time*, $\pi: M \times \boldsymbol{R}^2 \mapsto M \times \boldsymbol{R}$. *Among the totality of geodesics*, *only those whose arc-lengths are positive definite and are given by*

$$\mathrm{d}s^2 = c_1^2\, \mathrm{d}t^2 \tag{17}$$

correspond to natural motions; *condition* (17) *can be equivalently cast in the following integral form as a condition on the extra-coordinate* q^{N+1}:

$$q^{N+1} = \frac{c_1^2}{2} t + c_2^2 - \int_0^t \mathscr{L}\, \mathrm{d}\tau , \tag{18}$$

where c_1 *and* c_2 *are given real constants. Conversely*, *given a point* $P \in M \times \boldsymbol{R}$ *belonging to a trajectory of the system*, *and given two constants* c_1 *and* c_2, *the point* $P' = \pi^{-1}(P) \in M \times \boldsymbol{R}^2$, *with* q^{N+1} *given by* (18), *describes a geodesic curve in* $(M \times \boldsymbol{R}^2, g_E)$ *such that* $\mathrm{d}s^2 = c_1^2\, \mathrm{d}t^2$.

For the full proof, see Ref. [54]. Since the constant c_1 is arbitrary, we will always set $c_1^2 = 1$ in order that $\mathrm{d}s^2 = \mathrm{d}t^2$ on the physical geodesics.

From Eq. (16) follows that the explicit table of the components of the Eisenhart metric is

$$g_E = \begin{pmatrix} -2V(q) & 0 & \cdots & 0 & 1 \\ 0 & a_{11} & \cdots & a_{1N} & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & a_{N1} & \cdots & a_{NN} & 0 \\ 1 & 0 & \cdots & 0 & 0 \end{pmatrix}, \tag{19}$$

where a_{ij} is the kinetic energy metric. The non-vanishing Christoffel symbols, in the case $a_{ij} = \delta_{ij}$, are only

$$\Gamma^i_{00} = -\Gamma^{N+1}_{0i} = \partial_i V, \tag{20}$$

so that the geodesic equations (12) read

$$\frac{\mathrm{d}^2 q^0}{\mathrm{d}s^2} = 0, \tag{21a}$$

$$\frac{\mathrm{d}^2 q^i}{\mathrm{d}s^2} + \Gamma^i_{00}\frac{\mathrm{d}q^0}{\mathrm{d}s}\frac{\mathrm{d}q^0}{\mathrm{d}s} = 0, \tag{21b}$$

$$\frac{\mathrm{d}^2 q^{N+1}}{\mathrm{d}s^2} + \Gamma^{N+1}_{0i}\frac{\mathrm{d}q^0}{\mathrm{d}s}\frac{\mathrm{d}q^i}{\mathrm{d}s} = 0; \tag{21c}$$

using $\mathrm{d}s = \mathrm{d}t$ one obtains

$$\frac{\mathrm{d}^2 q^0}{\mathrm{d}t^2} = 0, \tag{22a}$$

$$\frac{\mathrm{d}^2 q^i}{\mathrm{d}t^2} = -\frac{\partial V}{\partial q_i}, \tag{22b}$$

$$\frac{\mathrm{d}^2 q^{N+1}}{\mathrm{d}t^2} = -\frac{\mathrm{d}\mathscr{L}}{\mathrm{d}t}. \tag{22c}$$

Eq. (22a) only states that $q^0 = t$, The N equations (22b) are Newton's equations, and Eq. (22c) is the differential version of Eq. (18).

The fact that in the framework of the Eisenhart metric the dynamics can be geometrized with an affine parametrization of the arc-length, i.e., $\mathrm{d}s = \mathrm{d}t$, will be extremely useful in the following, together with the remarkably simple curvature properties of the Eisenhart metric (see Section 3.3).

3.2. Curvature and stability

The geometrization of the dynamics is a natural framework for the study of the stability of the trajectories of a dynamical system, for it links the latter with the stability of the geodesics; the latter is *completely determined* by the curvature of the manifold, as shown below.

Studying the stability of the dynamics means determining the evolution of perturbations of a given trajectory. This implies that one should follow the evolution of the linearized (tangent) flow along the reference trajectory. For a Newtonian system, writing the perturbed trajectory as

$$\tilde{q}^i(t) = q^i(t) + \xi^i(t), \tag{23}$$

substituting this expression in the equations of motion

$$\ddot{q}^i = -\frac{\partial V(q)}{\partial q^i} \tag{24}$$

and retaining terms up to first order in the ξ's, one finds that the perturbation obeys the so-called *tangent dynamics equation* which reads as

$$\ddot{\xi}^i = -\left(\frac{\partial^2 V(q)}{\partial q^i \partial q^j}\right)_{q^i = q^i(t)} \xi^j. \tag{25}$$

This equation should be solved together with the dynamics in order to determine the stability or instability of the trajectory: when the norm of the perturbations grows exponentially, the trajectory is unstable, otherwise it is stable.

Let us now translate the stability problem into geometric language. By writing, in close analogy to what has been done above in the case of dynamical systems, a perturbed geodesic as

$$\tilde{q}^i(s) = q^i(s) + J^i(s) \tag{26}$$

and then inserting this expression in the equation for geodesics (12), one finds that the evolution of the perturbation vector J is given by the following equation:

$$\frac{\mathrm{D}^2 J^i}{\mathrm{d}s^2} + R^i_{jkl}\frac{\mathrm{d}q^j}{\mathrm{d}s} J^k \frac{\mathrm{d}q^l}{\mathrm{d}s} = 0, \tag{27}$$

where R^i_{jkl} are the components of the Riemann curvature tensor (see Eq. (A.39)). Eq. (27) is referred to as the Jacobi equation, and the tangent vector field J as the Jacobi field. This equation was first studied by Levi–Civita and is also often referred to as the equation of Jacobi and Levi–Civita. For a derivation we refer to Appendix A.4, where it is also shown that one can always assume that J is orthogonal to the velocity vector along the geodesic, $\dot{\gamma}$, i.e.,

$$\langle J, \dot{\gamma} \rangle = 0, \tag{28}$$

where $\langle \bullet, \bullet \rangle$ stands for the scalar product induced by the metric (see Eq. (A.20)). The remarkable fact is that the evolution of J – and then the stability or instability of the geodesic – is completely determined by the *curvature* of the manifold. Therefore, if the metric is induced by a physical system, as in the case of Jacobi or Eisenhart metrics, such an equation links the stability or instability of the trajectories to the curvature of the ambient manifold.

The subject of the next sections is precisely to exploit such a link in order to describe and understand the behaviour of those physical systems whose trajectories are mainly unstable.

However, before that, we have to give explicit expressions for the curvature of the mechanical manifolds, i.e., of those manifolds whose Riemannian structure is induced by the dynamics via the Jacobi or the Eisenhart metric.

3.3. Curvature of the mechanical manifolds

We already observed that the Jacobi metric is a conformal deformation of the kinetic-energy metric, whose components are given by the kinetic energy matrix a_{ij}. In the case of systems whose kinetic energy matrix is diagonal, this means that the Jacobi metric is conformally flat (see Appendix A.3). This greatly simplifies the computation of curvatures. It is convenient to define then a symmetric tensor C whose components are [11]

$$C_{ij} = \frac{N-2}{4(E-V)^2}\left[2(E-V)\partial_i\partial_j V + 3\partial_i V\partial_j V - \frac{\delta_{ij}}{2}|\nabla V|^2\right], \tag{29}$$

where V is the potential, E is the energy, and ∇ and $|\cdot|$ stand for the Euclidean gradient and norm, respectively. The curvature of (M_E, g_J) can be expressed through C. In fact, the components of the Riemann tensor are

$$R_{ijkm} = \frac{1}{N-2}[C_{jk}\delta_{im} - C_{jm}\delta_{ik} + C_{im}\delta_{jk} - C_{ik}\delta_{jm}]. \tag{30}$$

By contraction of the first and third indices, we obtain the Ricci tensor, whose components are (see Eq. (A.46))

$$R_{ij} = \frac{N-2}{4(E-V)^2}[2(E-V)\partial_i\partial_j V + 3\partial_i V\partial_j V] + \frac{\delta_{ij}}{4(E-V)^2}[2(E-V)\Delta V - (N-4)|\nabla V|^2] \tag{31}$$

and by a further contraction we obtain the scalar curvature (see Eq. (A.48))

$$\mathscr{R} = \frac{N-1}{4(E-V)^2}[2(E-V)\Delta V - (N-6)|\nabla V|^2]. \tag{32}$$

The curvature properties of the Eisenhart metric are much simpler than those of the Jacobi metric, and this is obviously a great advantage from a computational point of view. The only non-vanishing components of the curvature tensor are

$$R_{0i0j} = \partial_i\partial_j V, \tag{33}$$

hence the Ricci tensor has only one nonzero component

$$R_{00} = \Delta V \tag{34}$$

and the scalar curvature is identically vanishing,

$$\mathscr{R} = 0. \tag{35}$$

To summarize, we have shown that the dynamical trajectories of a Hamiltonian system of the form (1) can be seen as geodesics of the configuration space, or of an enlargement of it, once

a suitable metric is defined.[8] The general relationship which holds between dynamical and geometrical quantities regardless of the precise choice of the metric can be sketched as follows:

dynamics	*geometry*
(time) t	$\sim$ s *(arc-length)*
(potential energy) V	$\sim$ g *(metric)*
(forces) ∂V	$\sim$ Γ *(Christoffel symbols)*
(curvature of the potential) $\partial^2 V, (\partial V)^2$	$\sim$ R *(curvature of the manifold)*

(36)

In the case of the Eisenhart metric, all these relations are extremely simple (maybe as simple as possible). In fact, the physical time t can be chosen as equal to the arc-length s, the metric tensor g_E contains only the potential energy V, the non-vanishing Christoffel symbols Γ are equal to the forces ∂V, and the components of the Riemann curvature tensor R contain only the second derivatives of the potential energy, $\partial_i \partial_j V$.

We have also shown that the stability of the dynamical trajectories can be mapped onto the stability of the geodesics, which is completely determined by the curvature of the manifold. We will show in Section 4.1.3 that, in the case of the Eisenhart metric, as a consequence of its remarkably simple properties, also the relationship between the stability of the trajectories and the stability of the geodesics becomes as simple as possible, i.e., the Jacobi equation (27) becomes identical to the tangent dynamics equation (25).

4. Geometry and chaos

The purpose of the present section is to describe in some detail how it is possible, using the Jacobi equation as the main tool, to reach a twofold objective: first, to obtain a deeper understanding of the origin of chaos in Hamiltonian systems, and second, to obtain quantitative informations on the "strength" of chaos in these systems. Some basic concepts about Hamiltonian chaos and the definition of Lyapunov exponents are summarized in Appendix C.

4.1. Geometric approach to chaotic dynamics

A physical theory should provide a conceptual framework for modeling and understanding – at least at a qualitative level – the observed features of the system which is the object of the theory, and should also have a predictive content, i.e. should provide quantitative tools apt to compute, at least approximately, the outcomes of the experiments (no matter if it concerns laboratory experiments or numerical experiments performed on a computer). According to these requirements, a satisfactory theory of deterministic chaos is certainly still lacking. In fact, in both aspects the current theoretical approaches to chaos have some problems, especially if we consider the case of conservative flows, i.e., of the dynamics of conservative systems of ordinary differential equations.

[8] As already stated at the beginning of this section, there are many other possible choices for the ambient manifold and its metric: some other possible choices are described in Ref. [33].

To explain the origin of chaos in conservative dynamics one usually invokes the existence of invariant hyperbolic sets – or horseshoes – in phase space, like those generated by homoclinic intersections of perturbed separatrices (see Appendix C). In order to quantify the degree of instability of a trajectory or of a system we must instead resort to the notion of Lyapunov exponents. The Lyapunov exponents are asymptotic quantities and their relation with local properties of phase space, like horseshoes, is far from evident; nonetheless they provide the natural measure of the degree of chaos, measuring the typical time scales over which a trajectory looses the memory of its initial conditions. A rigorous definition of the existence of chaotic regions in the phase space of a system, based on the detection of horseshoes, does not provide any quantitative tool to measure chaos; on the other hand, Lyapunov exponents allow a very precise measure of chaos but give no information at all on the origin of such a chaotic behaviour. From a conceptual point of view this situation is far from being satisfactory, not to speak of the fact that the practical application of the methods to search horseshoes becomes extremely difficult as the number of degrees of freedom is large [55]. From the predictive point of view the situation is even worse, for no analytic method at all exists to compute Lyapunov exponents, at least in the case of flows of physical relevance. It is worth noticing that in a recent paper [56], Gozzi and Reuter have shown that one could build, in principle, a field-theoretic framework to compute Lyapunov exponents, but the practical application of such methods is still unclear. Needless to say, all the tools belonging to canonical perturbation theory, which have undergone remarkable developments in the last years [57], can hardly be used to compute quantities like Lyapunov exponents since in this framework one can only describe the regular, i.e., non-chaotic, features of phase space.

The geometric approach to dynamical instability allows a unification of the method to measure chaos with the explanation of its origin. In fact, the evolution of the field J given by the Jacobi equation (27) contains all the information needed to compute Lyapunov exponents, and makes us also recognize in the curvature properties of the ambient manifold the origin of chaotic dynamics.

Obviously, also this approach is far from being free of problems. For instance, the only case in which it is possible to rigorously prove that some definite curvature properties imply chaos in the geodesic flow, is the case of compact manifolds whose curvature is everywhere negative. In this case every point of the manifold is hyperbolic: in a sense this is the opposite limit to the integrable case. Though abstract and unphysical, such systems can help intuition. In a geodesic flow on a compact negatively curved manifold, the negative curvature forces nearby geodesics to separate exponentially, while the compactness ensures that such a separation does not reduce to a trivial "explosion" of the system and obliges the geodesics to fold. The joint action of stretching and folding is the essential ingredient of chaos.

Krylov tried to apply this framework to explain the origin of mixing in physical dynamical systems. Unfortunately, for many systems in which chaos is detected the curvatures are found mainly positive, and there are examples, for instance the Hénon–Heiles system – see Eq. (42) – geometrized with the Jacobi metric and the Fermi–Pasta–Ulam model – see Eq. (83a) – geometrized with the Eisenhart metric, where curvatures are always positive even in the presence of fully developed chaos. Hence even positive curvature must be able to produce chaos.

Only recently an example has been found of a compact surface with positive curvature, where the presence of chaotic regions coexisting with regular ones can be rigorously proved [58], and this provides mathematical support for the available numerical evidence that negative curvature is not necessary at all to have chaos in a geodesic flow [22,24,31]. What then *is* the crucial feature of the

curvature which is required to produce chaos? There is not yet a definite answer – at least on rigorous grounds – to this question. Nevertheless it is sure that, if positive, curvature must be non-constant in order to originate instability, and we shall see that the *curvature fluctuations* along the geodesics can be responsible for the appearance of an instability through a mechanism very close to parametric instability.

The advantages of the geometric approach to chaos are not only conceptual: also on predictive grounds this framework proves very useful. For, starting from the Jacobi equation, it is possible to obtain an effective stability equation which allows one to obtain an analytic estimate of the largest Lyapunov exponent in the thermodynamic limit [25,31]. Such an estimate turns out to be in very good agreement with the results of numerical simulations for a number of systems (see Section 4.2). In order to understand the derivation of such an effective stability equation, let us investigate in greater detail the relation between stability and curvature which was introduced in the last section.

4.1.1. Geometric origin of Hamiltonian chaos

Let us consider an N-dimensional Riemannian (or pseudo-Riemannian) manifold (M, g) and a local coordinate frame with coordinates $(q^1, \dots, q^N)$.

We already observed that the evolution of the Jacobi field J, which contains the whole information on the stability of the geodesic flow, is completely determined by the curvature tensor R through the Jacobi equation (27). Unfortunately, the number of independent components of the tensor R is $\mathcal{O}(N^4)$ – even if this number can be considerably reduced by symmetry considerations – so that Eq. (27) becomes rather untractable already at fairly small dimensions.

Nevertheless, there is a particular case in which the Jacobi equation has a remarkably simple form: the case of *isotropic* manifolds (see Appendix A.3) where Eq. (27) becomes

$$\frac{\mathrm{D}^2 J^i}{\mathrm{d}s^2} + K J^i = 0\,, \tag{37}$$

where K is the constant sectional curvature of the manifold (see Eq. (A.42)). Choosing a geodesic frame, i.e. an orthonormal frame transported along the geodesic, covariant derivatives become ordinary derivatives, i.e., $\mathrm{D}/\mathrm{d}s \equiv \mathrm{d}/\mathrm{d}s$, so that the solution of Eq. (37), with initial condition $J(0) = 0$ and $\mathrm{d}J(0)/\mathrm{d}s = w(0)$, is

$$J(s) = \begin{cases} \dfrac{w(s)}{\sqrt{K}} \sin(\sqrt{K}s) & (k > 0)\,, \\ s w(s) & (K = 0)\,, \\ \dfrac{w(s)}{\sqrt{-K}} \sinh(\sqrt{-K}\,s) & (K < 0)\,. \end{cases} \tag{38}$$

The geodesic flow is unstable only if $K < 0$, and in this case the instability exponent is just $\sqrt{-K}$.

As long as the curvatures are negative, the geodesic flow is unstable even if the manifold is no longer isotropic, and by means of the so-called "comparison theorems" (mainly Rauch's theorem, see e.g. [59]) it is possible to prove that the instability exponent is greater or equal to $(-\max_M(K))^{1/2}$ [7]. On the contrary, no exact results of general validity have yet been found for

the dynamics of geodesic flows on manifolds whose curvature is neither constant nor everywhere negative.

Eq. (37) is valid only if K is constant. Nevertheless, in the case in which $\dim M = 2$ (surfaces), the Jacobi equation – again written in a geodesic reference frame for the sake of simplicity – takes a form very close to that for isotropic manifolds,

$$\frac{\mathrm{d}^2 J}{\mathrm{d}s^2} + K(s)J = 0\,, \tag{39}$$

where

$$K(s) = \tfrac{1}{2}\mathscr{R}(s) \tag{40}$$

and, contrary to Eq. (37), it is no longer a constant. With $\mathscr{R}(s)$ we denote the scalar curvature of the manifold at the point $P = \gamma(s)$ (see Eq. (A.45)). If the geodesics are unstable, Eq. (39) has exponentially growing solutions. As far as we know [60], the solutions of (39) can exhibit an exponentially growing envelope in two cases:

(a) the curvature $K(s)$ takes negative values;
(b) the curvature $K(s)$, though mainly or even exclusively positive, fluctuates in such a way that it triggers a sort of *parametric instability* mechanism.

In the first case, the mechanism that is at the origin of the instability of the geodesics is the one usually considered in ergodic theory [7]. But in the second case a new mechanism of instability, which does not require the presence of negatively curved regions on the manifold, shows up: the fluctuations of the curvature along the geodesic make the geodesic unstable.

Let us now turn to physics, i.e., to the case of a mechanical manifold: in the case of the Jacobi metric with $N = 2$ the scalar curvature written in standard (Lagrangian) coordinates reads as

$$\mathscr{R} = \frac{(\nabla V)^2}{(E - V)^3} + \frac{\Delta V}{(E - V)^2}\,, \tag{41}$$

where ∇ and Δ stand, respectively, for the Euclidean gradient and Laplacian operators. Hence we can have $\mathscr{R} < 0$ only if $\Delta V < 0$, i.e., for stable physical potentials, when the potential has inflection points. In these cases Krylov's idea can work – even if in the high-dimensional case this becomes very complicated – and we may have dynamical chaos induced by negative curvatures of the manifold. Indeed Krylov was mainly concerned with weakly non-ideal gases, or in general dilute systems, where for typical interatomic interactions $\Delta V < 0$ so that the curvatures can be negative (see Krylov's Ph.D. Thesis in Ref. [5]).

We will now show one example in which, though chaos is present, curvatures are positive. Let us consider the Hénon–Heiles model [61], whose Hamiltonian is

$$\mathscr{H} = \tfrac{1}{2}(p_x^2 + p_y^2) + \tfrac{1}{2}(x^2 + y^2) + x^2 y - \tfrac{1}{3}y^3\,. \tag{42}$$

This model was introduced in an astrophysical framework to study the motion of a star in an axially symmetric galaxy, but it can also be regarded as a model of a triatomic molecule (after having used translational symmetry to eliminate the center-of-mass coordinate) [62]. The Hénon–Heiles model is a cornerstone in the study of Hamiltonian chaos: it was the first physical model for which chaos was found and where a transition from a mainly regular to a mainly chaotic

phase space was identified under a variation of the energy. In this model, Eq. (39) is exact, but $\mathscr{R} > 0$ everywhere. Hence chaos in this system cannot come from any negative curvature in the associated mechanical (Jacobi) manifold. As we shall see later on (see e.g. Section 4.2), the absence of negative curvatures in the associated mechanical manifolds is not a peculiarity of this model, for it is shared with many systems of interest for field theory and condensed matter physics which have chaotic trajectories. In particular, all the systems that in the low-energy limit behave as a collection of harmonic oscillators do belong to this class.

In these cases the second of the previously discussed instability mechanisms, the one mentioned in item (b), comes in: curvature fluctuations may induce chaos through parametric instability. The latter is a well-known feature of differential equations whose parameters are time-dependent. The classical example (see e.g. Arnol'd's book [2]) is the mathematical swing, i.e. a pendulum, initially at rest, whose length is modulated in time. If the modulation contains frequencies resonating with the free pendulum's fundamental frequency, the stable equilibrium position gets unstable and the swing starts to oscillate with growing amplitude. In Eq. (39), $\sqrt{K(s)}$ and s play the role of a frequency and of time, respectively, so that this equation can be thought of as the equation of motion of a harmonic oscillator with time-dependent frequency, often referred to as a (generalized) Hill's equation [63]. By expanding $K(s)$ in a Fourier series we get

$$K(s) = K_0 + \sum_{n=1}^{\infty} [a_n \cos(n\omega s) + b_n \sin(n\omega s)], \tag{43}$$

where $\omega = 2\pi/L$ and L is the length of the geodesic. The presence of resonances between the average frequency $\sqrt{K_0}$ and the frequency in some term in expansion (43) eventually forces an exponential growth of the solutions of the equation. In the simplest case, in which only one coefficient of series (43), say a_1, is non-vanishing, the equation is called the Mathieu equation and it is possible to compute analytically both the bounds of the instability regions in the parameter space and the actual value of the instability exponent [63]. At variance with the Mathieu case, in the general case, where a large number of coefficients of the Fourier decomposition of $K(s)$ is non-zero, it is much more difficult to do something similar. Hence there is not yet any rigorous proof of the fact that this kind of parametric instability is the mechanism that produces chaos in Hamiltonian dynamical systems – in the two-degrees-of-freedom case or in the general case – and this still remains a conjecture. Nevertheless, such a conjecture is strongly supported by at least two facts. First of all, in recent papers [30,27] it has been shown that the solutions of the Jacobi equation (39) for the Hénon–Heiles model and for a model of quartic coupled oscillators show an oscillatory behaviour with an exponentially growing envelope – which is precisely what one expects from parametric instability – in the chaotic regions, while the oscillations are bounded in the regular regions. Second, also in high-dimensional flows the components of the Jacobi field J oscillate with an exponentially growing amplitude as long as the system is non-integrable, whereas they exhibit only bounded oscillations for integrable systems. Moreover, in the high-dimensional case (i.e., for systems with a large number of degrees of freedom) it is possible to establish a *quantitative* link between the largest Lyapunov exponent and the curvature fluctuations. In fact, as we shall see in the following, in the high-dimensional case it is possibile to write down, under suitable approximations, an effective stability equation which looks very similar to Eq. (39), but where the squared frequency $K(s)$ is a stochastic process, and, through this equation, it is possible to give an analytical

estimate of the largest Lyapunov exponent. Since, from now on, we are going to consider only the largest Lyapunov exponent, the latter will be referred to as just the Lyapunov exponent.

4.1.2. Effective stability equation in the high-dimensional case

Let us now study the problem of the stability of the geodesics in manifolds whose dimension N is large: according to the correspondence between geometry and dynamics introduced in Section 3, we are considering a system with a large number N of degrees of freedom.

Our starting point is the Jacobi equation (27). Our aim is to derive from it an effective stability equation which no longer depends on the dynamics, i.e., on the evolution of the particular geodesic that we are following, but only on the average curvature properties of the manifold. To do that, we need some assumptions and approximations which are not valid in general but which are very reasonable in the case of large-N mechanical manifolds. For the sake of clarity, we first summarize the assumptions and approximations leading to our final result, and later on we discuss them more thoroughly. Further details can be found in the papers where this approach was originally put forward [25,31]:

0. We assume that the evolution of a generic geodesic is chaotic. This assumption is reasonable in the case of a manifold whose geodesics are the trajectories of a generic Hamiltonian system with a large number of degrees of freedom N, for in this case the overwhelming majority of the trajectories will be chaotic. This bears a certain similarity to Gallavotti and Cohen's "chaotic hypothesis" [64].
1. We assume that the manifold is *quasi-isotropic*. Loosely speaking, this assumption means that the manifold can be regarded somehow as a locally deformed constant-curvature manifold. However, we will give this assumption a precise formulation later, in Eqs. (49). This approximation allows us to get rid of the dependence of the Jacobi equation (27) on the full Riemann curvature tensor by replacing it with an effective sectional curvature $\mathscr{K}(s)$ along the geodesic; moreover, the Jacobi equation becomes diagonal.
2. To get rid of the dependence of the effective sectional curvature $\mathscr{K}(s)$ on the dynamics, i.e., on the evolution of the geodesic, we model $\mathscr{K}(s)$ with a stochastic process. This assumption is motivated by Assumption 0 above. Moreover, as long as we consider a high-dimensional mechanical manifold associated to a Hamiltonian flow with N degrees of freedom and we are eventually interested in taking the thermodynamic limit $N \to \infty$, the sectional curvature is formed by adding up many independent terms, so that invoking a central-limit-theorem-like argument, $\mathscr{K}(s)$ is expected to behave, in first approximation, as a Gaussian stochastic process.
3. We assume that the statistics of the effective sectional curvature $\mathscr{K}$ is the same as that of the Ricci curvature K_R, which is a suitably averaged sectional curvature (see Eq. (A.44)). Such an assumption is consistent with Assumption 1 above, for in a constant-curvature manifold the sectional curvature equals the Ricci curvature times a constant, and allows us to compute the mean and the variance of the stochastic process introduced in Assumption 2 in terms of the average and the variance of K_R along a generic geodesic.
4. The last step, which completely decouples the problem of the stability of the geodesics from the evolution of the geodesics themselves, consists in replacing the (proper) time averages of the Ricci curvature with static averages computed with a suitable probability measure μ. If the

manifold is a mechanical manifold, the natural choice for μ is the microcanonical measure. Again this assumption is reasonable if Assumption 0 is valid.

After these steps, we end up with an effective stability equation which no longer depends on the evolution of the geodesics, but only on the average and fluctuations of the Ricci curvature of the manifold.

Let us now discuss more thoroughly the above-sketched procedure. For that, it is convenient to introduce the Weyl projective tensor W, whose components are given by [65]

$$W^i_{jkl} = R^i_{jkl} - \frac{1}{N-1}(R_{jl}\delta^i_k - R_{jk}\delta^i_l), \tag{44}$$

where $R_{ij} = R^m_{imj}$ are the components of the Ricci curvature tensor (see Eq. (A.46)). Weyl's projective tensor measures the deviation from isotropy of a given manifold, since it vanishes identically for an isotropic manifold. Then we can reformulate the Jacobi equation (27) in the following form [31]:

$$\frac{\mathrm{D}^2 J^i}{\mathrm{d}s^2} + \frac{1}{N-1}R_{jk}\frac{\mathrm{d}q^j}{\mathrm{d}s}\frac{\mathrm{d}q^k}{\mathrm{d}s}J^i - \frac{1}{N-1}R_{jk}\frac{\mathrm{d}q^j}{\mathrm{d}s}J^k\frac{\mathrm{d}q^i}{\mathrm{d}s} + W^i_{jkl}\frac{\mathrm{d}q^j}{\mathrm{d}s}J^k\frac{\mathrm{d}q^l}{\mathrm{d}s} = 0. \tag{45}$$

For an isotropic manifold the third term in Eq. (45) vanishes because $R_{jk} = K g_{jk}$ (see Eq. (A.50)) so that $R_{jk}\dot{q}^j J^k = K\langle\dot{\gamma}, J\rangle$, and $\langle\dot{\gamma}, J\rangle = 0$ (see Eq. (28)). Thus, for an isotropic manifold Eq. (45) collapses to Eq. (37), in fact the second term is nothing but $K J^i$. When the manifold is not isotropic, we see that Eq. (45) retains the structure of Eq. (37) up to its second term, since the coefficient of J^i is still a scalar. This coefficient has now the meaning of a sectional curvature averaged, at any given point, over the $N-1$ independent directions orthogonal to $\dot{\gamma}$, the velocity vector of the geodesic. However, such a mean sectional curvature is no longer constant along the geodesic $\gamma(s)$, and is just the Ricci curvature K_R divided by $N-1$ (see Eq. (A.47)). The fourth term of (45) accounts for the local degree of anisotropy of the ambient manifold.

Let us now consider a geodesic frame: in this case Eq. (45) can be rewritten as

$$\frac{\mathrm{d}^2 J^i}{\mathrm{d}s^2} + k_R(s)\,J^i - r^i_j(s)\,J^j + w^i_j(s)\,J^j = 0, \tag{46}$$

where, to ease the notation, we have put

$$k_R(s) = \frac{K_R}{N-1} = \frac{1}{N-1}R_{jk}\frac{\mathrm{d}q^j}{\mathrm{d}s}\frac{\mathrm{d}q^k}{\mathrm{d}s}, \tag{47a}$$

$$r^i_j(s) = \frac{1}{N-1}R_{jk}\frac{\mathrm{d}q^k}{\mathrm{d}s}\frac{\mathrm{d}q^i}{\mathrm{d}s}, \tag{47b}$$

$$w^i_j(s) = W^i_{kjl}\frac{\mathrm{d}q^k}{\mathrm{d}s}\frac{\mathrm{d}q^l}{\mathrm{d}s}. \tag{47c}$$

Being a scalar quantity, the value of k_R is independent of the coordinate system. Now let us formulate our Assumption 1, namely, that the manifold is quasi-isotropic, in a more precise way.

To do that, we recall that (see Eqs. (A.49) and (A.50)) if and only if the manifold is isotropic, i.e., has constant curvature, the Riemann curvature tensor and the Ricci tensor can be written in the remarkably simple forms, i.e.,

$$R_{ijkl} = K(g_{ik}g_{jl} - g_{il}g_{jk}) \tag{48a}$$

and

$$R_{ij} = K\, g_{ij}\,, \tag{48b}$$

where K is a scalar constant, the sectional curvature of the manifold. The precise formulation of Assumption 1 is now that along a generic geodesic the Riemann curvature tensor and the Ricci tensor retain the same functional form as in case (48), i.e., that

$$R_{ijkl} \approx \mathscr{K}(s)\,(g_{ik}g_{jl} - g_{il}g_{jk}) \tag{49a}$$

and

$$R_{ij} \approx \mathscr{K}(s)\, g_{ij}\,, \tag{49b}$$

where $\mathscr{K}(s)$, which is no longer a constant, is an effective sectional curvature. In the general case we are not able to give a rigorous explicit expression for $\mathscr{K}(s)$, because the functional dependence postulated in Eqs. (49) holds true only for constant-curvature manifolds. However, the effective curvature $\mathscr{K}(s)$ is expected to be essentially the sectional curvature $K(\dot\gamma, J)$ (see Eq. (A.44)) measured along the geodesic in the directions of the velocity vector $\dot\gamma = \mathrm{d}q/\mathrm{d}s$ and of the Jacobi vector J.

Combining Eqs. (47b) and (49b), and recalling that the vector J is orthogonal to the velocity of the geodesic, i.e., $g_{ij}(\mathrm{d}q^i/\mathrm{d}s)J^j = 0$, we find that the third term in Eq. (46), $-r^i_j J^j$, vanishes as in the isotropic case. Now we combine Eqs. (44) and (49a) to obtain

$$W^i_{jkl} \approx \mathscr{K}(s)(\delta^i_j g_{kl} - \delta^i_l g_{kj}) - \frac{1}{N-1}(R_{jl}\delta^i_k - R_{jk}\delta^i_l)\,, \tag{50}$$

so that Eq. (47c) can be rewritten as

$$w^i_j \approx \mathscr{K}(s)\delta^i_j - k_R(s)\delta^i_j - \mathscr{K}(s)\frac{N-2}{N-1}\frac{\mathrm{d}q^i}{\mathrm{d}s}g_{kj}\frac{\mathrm{d}q^k}{\mathrm{d}s}\,, \tag{51}$$

where we have used the definition of k_R given in Eq. (47a) and approximation (49b) for the Ricci tensor. Let us now insert Eq. (51) into Eq. (46): the last term of Eq. (51) vanishes after having been multiplied by J^j and summed over j, because J and $\mathrm{d}q/\mathrm{d}s$ are orthogonal, and the term $k_R(s)J^i$ is cancelled by the term $-k_R(s)J^i$ coming from Eq. (51), so that Eq. (46) is finally rewritten as

$$\frac{\mathrm{d}^2 J^i}{\mathrm{d}s^2} + \mathscr{K}(s)\, J^i = 0\,. \tag{52}$$

Eq. (52) is now diagonal. However, in order to use it, we should know the values of $\mathscr{K}(s)$ along the geodesic. Here, Assumptions 2 and 3 come into play: we replace $\mathscr{K}(s)$ with a stochastic Gaussian process, and we assume that its probability distribution is the same as that of the Ricci curvature,

$$\mathscr{P}(\mathscr{K}) \simeq \mathscr{P}(K_R)\,. \tag{53}$$

Such an assumption is consistent with our Assumption 1, because for an isotropic manifold the sectional curvature is identical to the Ricci curvature divided by $N-1$, so that, if the manifold is quasi-isotropic, it is natural to assume that the probability distributions of the sectional curvature and of the Ricci curvature are similar. Moreover, such an assumption is also the only easy one, because we are able to compute, under some further assumptions, the probability distribution of K_R, but we do not know anything about $\mathscr{K}$.

To be consistent with the definition of the sectional and the Ricci curvatures (see Eq. (A.51)), the following relations are assumed to hold for the first two cumulants of (53):

$$\langle \mathscr{K}(s)\rangle_s \simeq \frac{1}{N-1}\langle K_R(s)\rangle_s \equiv \langle k_R(s)\rangle_s\,, \tag{54a}$$

$$\langle [\mathscr{K}(s)-\bar{\mathscr{K}}]^2\rangle_s \simeq \frac{1}{N-1}\langle [K_R(s)-\langle K_R\rangle_s]^2\rangle_s \equiv \langle \delta^2 k_R\rangle_s\,, \tag{54b}$$

where $\langle\cdot\rangle_s$ stands for a proper-time average along a geodesic $\gamma(s)$. In general, the probability distributions (53) will not be Gaussian, i.e., other cumulants in addition to the first two will be non-vanishing. However, we already observed that since for a large system K_R is obtained by summing a large number of independent components, it is reasonable to assume that a sort of central limit theorem holds and that a Gaussian approximation is sufficient.

Our approximation for the effective sectional curvature $\mathscr{K}(s)$ is then the stochastic process

$$\mathscr{K}(s) \simeq \langle k_R(s)\rangle_s + \langle \delta^2 k_R\rangle_s^{1/2}\eta(s)\,, \tag{55}$$

where $\eta(s)$ is a random Gaussian process with zero mean and unit variance.

Finally, in order to completely decouple the stability equation from the dynamics, we use Assumption 4 and we replace time averages with static averages computed with a suitable measure μ. If the manifold is a mechanical manifold the geodesics are the natural motions of the systems, and a natural choice for μ is then the microcanonical ensemble, so that Eq. (55) becomes

$$\mathscr{K}(s) \simeq \langle k_R(s)\rangle_\mu + \langle \delta^2 k_R\rangle_\mu^{1/2}\,\eta(s)\,. \tag{56}$$

Our final effective stability equation is then

$$\frac{\mathrm{d}^2\psi}{\mathrm{d}s^2} + \langle k_R\rangle_\mu\,\psi + \langle \delta^2 k_R\rangle_\mu^{1/2}\,\eta(s)\,\psi = 0\,, \tag{57}$$

where ψ stands for *any* of the components of J, since all of them now obey the *same* effective equation of motion.

Eq. (57) implies that, if the manifold is a mechanical manifold, the growth-rate of ψ gives the dynamical instability exponent in our Riemannian framework. Eq. (57) is a scalar equation which, *independently of the knowledge of the dynamics*, provides a measure of the degree of instability of the dynamics itself through the behaviour of $\psi(s)$. The peculiar properties of a given Hamiltonian system enter Eq. (57) only through the global geometric properties $\langle k_R\rangle_\mu$ and $\langle \delta^2 k_R\rangle_\mu$ of the ambient Riemannian manifold (whose geodesics are natural motions) and are sufficient, as long as our Assumptions 1–4 hold, to determine the average degree of chaoticity of the dynamics. Moreover, $\langle k_R\rangle_\mu$ and $\langle \delta^2 k_R\rangle_\mu$ are microcanonical averages, so that they are functions of the

energy E of the system, or of the energy per degree of freedom $\varepsilon = E/N$ which is the relevant parameter as $N \to \infty$. Thus from (57) we can obtain the energy dependence of the geometric instability exponent.

Within the validity of our Assumptions 1–4, transforming the Jacobi equation (27) into Eq. (57), the original complexity of the Jacobi equation has been considerably reduced: from a tensor equation we have obtained an effective scalar equation formally representing the equation of motion of a stochastic oscillator. In fact, Eq. (57), with a self-evident notation, is of the form

$$\frac{\mathrm{d}^2\psi}{\mathrm{d}s^2} + k(s)\,\psi = 0\,, \tag{58}$$

where $k(s)$, the squared frequency, is a Gaussian stochastic process.

Moreover, such an equation admits a very suggestive geometric interpretation, since it is a scalar equation, i.e., it is formally the Jacobi equation on a two-dimensional manifold whose Gaussian curvature is given, along a geodesic, by the random process $k(s)$ and can be regarded as an "effective" low-dimensional manifold approximating the "true" high-dimensional manifold where the dynamics of the geodesic flow takes place. This is the real geometrical content of our quasi-isotropy hypothesis. Hence, the average global curvature properties $\langle k_R \rangle_\mu$ and $\langle \delta^2 k_R \rangle_\mu$, in addition to being the ingredients for a geometric computation of the instability exponent, convey also information on the geometric structure of this effective manifold. Thus, we expect that it will be possible to gain some insight in the global properties of the dynamics by simply studying the behaviour of these average curvature properties as the energy is varied.

4.1.3. A geometric formula for the Lyapunov exponent

Let us now study the properties of the solutions of Eq. (58) in order to obtain an analytic estimate for the Lyapunov exponent. The derivation of the stochastic oscillator equation does not depend on a particular choice of the metric; within the approximations discussed above, Eq. (58) holds regardless of the choice of the metric. However, to make explicit the connection between the solutions of Eq. (58) and the stability of a dynamical system, one has to choose a particular metric; in the case of Hamiltonian systems of the form (1), the choice of the Eisenhart metric is the simplest one.

For this reason, we shall from now on restrict ourselves to standard Hamiltonian systems with a diagonal kinetic energy matrix, i.e., $a_{ij} = \delta_{ij}$, choosing as ambient manifold for the geometrization of the dynamics the enlarged configuration space–time equipped with the Eisenhart metric (16). The case of the Jacobi metric is discussed in Ref. [31].

We will proceed as follows. (i) We will show that in the present case the Jacobi equation (27) is *equal* to the tangent dynamics equation (25). (ii) We will replace the arc-length s with the time t and we will explicitly compute the average and the fluctuations of the Ricci curvature along a geodesic in terms of dynamical observables, so that the (static) probability distribution of the stochastic process $k(t)$ which models the effective sectional curvature is defined. (iii) We will give an estimate for the time correlation function of the process $k(t)$. (iv) We will solve the stochastic oscillator equation, obtaining an analytical formula for the Lyapunov exponent.

Let us now consider item (i). As a consequence of the simple structure of the curvature tensor for the Eisenhart metric (see Section 3.3), the Jacobi equation (27) takes the form (we recall that the manifold has now dimension $N+2$; all the indices $i,j,k,\ldots$ run from 1 to N)

$$\frac{\mathrm{D}^2 J^0}{\mathrm{d}s^2} + R^0_{i0j}\frac{\mathrm{d}q^i}{\mathrm{d}s}J^0\frac{\mathrm{d}q^j}{\mathrm{d}s} + R^0_{0ij}\frac{\mathrm{d}q^0}{\mathrm{d}s}J^i\frac{\mathrm{d}q^j}{\mathrm{d}s} = 0\,, \tag{59a}$$

$$\frac{\mathrm{D}^2 J^i}{\mathrm{d}s^2} + R^i_{0j0}\left(\frac{\mathrm{d}q^0}{\mathrm{d}s}\right)^2 J^j + R^i_{00j}\frac{\mathrm{d}q^0}{\mathrm{d}s}J^0\frac{\mathrm{d}q^j}{\mathrm{d}s} + R^i_{j00}\frac{\mathrm{d}q^j}{\mathrm{d}s}J^0\frac{\mathrm{d}q^0}{\mathrm{d}s} = 0\,, \tag{59b}$$

$$\frac{\mathrm{D}^2 J^{N+1}}{\mathrm{d}s^2} + R^{N+1}_{i0j}\frac{\mathrm{d}q^i}{\mathrm{d}s}J^0\frac{\mathrm{d}q^j}{\mathrm{d}s} + R^{N+1}_{ij0}\frac{\mathrm{d}q^i}{\mathrm{d}s}J^j\frac{\mathrm{d}q^0}{\mathrm{d}s} = 0\,, \tag{59c}$$

where, for the sake of clarity, we have written out Eq. (59) separately for the 0, the $i=1,\ldots,N$, and the $N+1$ components, respectively. As $\Gamma^0_{ij}=0$ (see Eq. (20)) we obtain, from the definition of covariant derivative (see Eq. (A.32)), $\mathrm{D}J^0/\mathrm{d}s = \mathrm{d}J^0/\mathrm{d}s$, and, as $R^0_{ijk}=0$ (see Section 3.3), we find that Eq. (59a) becomes

$$\frac{\mathrm{d}^2 J^0}{\mathrm{d}s^2} = 0\,, \tag{60}$$

so that J^0 does not accelerate and, without loss of generality, we can set $\mathrm{d}J^0/\mathrm{d}s|_{s=0} = J^0(0) = 0$. Combining the latter result with the definition of covariant derivative we obtain

$$\frac{\mathrm{D}J^i}{\mathrm{d}s} = \frac{\mathrm{d}J^i}{\mathrm{d}s} + \Gamma^i_{0k}\frac{\mathrm{d}q^0}{\mathrm{d}s}J^k \tag{61}$$

and using $\mathrm{d}q^0/\mathrm{d}s = 0$ we finally get

$$\frac{\mathrm{D}^2 J^i}{\mathrm{d}s^2} = \frac{\mathrm{d}^2 J^i}{\mathrm{d}s^2}\,, \tag{62}$$

so that Eq. (59b) gives, for the projection in configuration space of the separation vector,

$$\frac{\mathrm{d}^2 J^i}{\mathrm{d}s^2} + \frac{\partial^2 V}{\partial q_i \partial q_k}\left(\frac{\mathrm{d}q^0}{\mathrm{d}s}\right)^2 J_k = 0\,. \tag{63}$$

Eq. (59c) describes the evolution of J^{N+1}, which, however, does not contribute to the norm of J because $g_{N+1N+1}=0$, so we can disregard it.

Along the physical geodesics of g_E, $\mathrm{d}s^2 = (\mathrm{d}q^0)^2 = \mathrm{d}t^2$, so that Eq. (63) is exactly the usual tangent dynamics equation (25), provided the identification $\xi = J$ is made. This clarifies then the relationship between the geometric description of the instability of a geodesic flow and the conventional description of dynamical instability. We stress that from a formal viewpoint this is a peculiarity of the Eisenhart metric; nevertheless, the physical content of this result is valid independently of the metric used, as long as the identification between trajectories and geodesics holds true. For, in recent papers [30,27] it has been found that using the Jacobi metric the solutions of the Jacobi equation and of those of the tangent dynamics equation – which in this case are two well-distinct equations – look strikingly similar.

We now turn to item (ii). The Ricci curvature is obtained saturating the Ricci tensor with the components of the velocity vector $\mathrm{d}q/\mathrm{d}s$ (see Eq. (A.47)). In the present case, the only non-vanishing component of the Ricci tensor is $R_{00} = \Delta V$ (see Eq. (34)), so that the dynamical observable which corresponds to the Ricci curvature along a geodesic depends only on the coordinates and not on the velocities and reads

$$K_R(q) = \Delta V\,, \tag{64}$$

where we have used that, along a physical geodesic, $(\mathrm{d}q^0)^2 = \mathrm{d}t^2 - \mathrm{d}s^2$. Using again this result we replace the arc-length s along the geodesic with the physical time t, and the stochastic oscillator equation (58) can be written

$$\frac{\mathrm{d}^2\psi}{\mathrm{d}t^2} + k(t)\psi = 0\,, \tag{65}$$

where mean and variance of $k(t)$ are given by

$$k_0 \equiv \langle k_R \rangle_\mu = \frac{1}{N}\langle \Delta V \rangle_\mu\,, \tag{66a}$$

$$\sigma_k^2 \equiv \langle \delta^2 k_R \rangle_\mu = \frac{1}{N}(\langle (\Delta V)^2 \rangle_\mu - \langle \Delta V \rangle_\mu^2)\,. \tag{66b}$$

Since we are considering systems with large N – eventually taking the limit $N \to \infty$ – we replaced $N - 1$ with N in Eqs. (66).

We consider now item (iii). The process $k(t)$ is not completely defined unless its time correlation function,

$$\Gamma_k(t_1, t_2) = \langle k(t_1)k(t_2)\rangle - \langle k(t_1)\rangle\langle k(t_2)\rangle\,, \tag{67}$$

is given. The simplest choice is to assume that $k(t)$ is a stationary and δ-correlated process, so that

$$\Gamma_k(t_1, t_2) = \Gamma_k(|t_2 - t_1|) = \Gamma_k(t) = \tau\sigma_k^2\delta(t)\,, \tag{68}$$

where τ is the characteristic correlation time scale of the process.

Before we can actually solve Eq. (65), we have then to give an explicit expression for τ. To do that, first we will show how two independent characteristic correlation time scales, which will be referred to as τ_1 and τ_2, respectively, can be defined, then we will estimate τ by combining these two time scales.

A first time scale, which we will refer to as τ_1, is associated to the time needed to cover the average distance between two successive *conjugate points* along a geodesic. Conjugate points [59] are the points where the Jacobi vector field vanishes. As long as the curvature is positive and its fluctuations are small compared to the average, two nearby geodesics will remain close to each other until a conjugate point is reached. At each crossing of a conjugate point the Jacobi vector field increases as if the geodesics were kicked (this is what happens when parametric instability is active). Thus the average distance between conjugate points provides a relevant correlation time scale. It can be proved that [59,66] if the sectional curvature K is bounded as $0 < L \leq K \leq H$, then

the distance d between two successive conjugate points is bounded by $d > \pi/2\sqrt{H}$. The upper bound H of the curvature can then be approximated in our framework by

$$H \simeq k_0 + \sigma_k \,, \tag{69}$$

so that we can define τ_1 as (remember that $\mathrm{d}t = \mathrm{d}s$)

$$\tau_1 = d_1 = \frac{\pi}{2\sqrt{k_0 + \sigma_k}} \,. \tag{70}$$

This time scale is expected to be the most relevant only as long as the curvature is positive and the fluctuations are small, compared to the average.

Another time scale, referred to as τ_2, is related to the local curvature fluctuations. These will be felt on a length scale of the order of, at least, $l = 1/\sqrt{\sigma_k}$ (the average fluctuation of curvature radius). The scale l is expected to be the relevant one when the fluctuations are of the same order of magnitude as the average curvature. Locally, the metric of a manifold can be approximated by [59]

$$g_{ik} \simeq \delta_{ik} - \tfrac{1}{6} R_{ikjl} u^i u^k \,, \tag{71}$$

where the u^i are the components of the displacements from the point around which we are approximating the metric. When the sectional curvature is positive (resp. negative), lengths and time intervals – on a scale l – are enlarged (resp. shortened) by a factor $(l^2K/6)$, so that the period $2\pi/\sqrt{k_0}$ has a fluctuation amplitude d_2 given by $d_2 = (l^2K/6)2\pi/\sqrt{k_0}$; replacing K by the most probable value k_0 one gets

$$\tau_2 = d_2 = \frac{l^2 k_0}{6} \frac{2\pi}{\sqrt{k_0}} \simeq \frac{k_0^{1/2}}{\sigma_k} \,. \tag{72}$$

Finally τ in Eq. (68) is obtained by combining τ_1 with τ_2 as follows:

$$\tau^{-1} = \tau_1^{-1} + \tau_2^{-1} \,. \tag{73}$$

The present definition of τ is obviously by no means a direct consequence of any theoretical result, but only a rough, physically based estimate. Such an estimate might well be improved independently of the general geometric framework.

Now that all the quantities entering Eq. (65) have been fully defined, we can turn to item (iv), i.e., to the solution of Eq. (65). Whenever $k(t)$ has a non-vanishing stochastic component, any solution $\psi(t)$ has an exponentially growing envelope [67] whose growth-rate provides a measure of the degree of instability. How can one relate such a growth-rate with the Lyapunov exponent of the physical system? Let us recall that, for a standard Hamiltonian system of the form (1), the Lyapunov exponent can be computed as the following limit (see Eq. (C.12)):

$$\lambda = \lim_{t\to\infty} \frac{1}{2t} \log \frac{\xi_1^2(t) + \cdots + \xi_N^2(t) + \dot{\xi}_1^2(t) + \cdots + \dot{\xi}_N^2(t)}{\xi_1^2(0) + \cdots + \xi_N^2(0) + \dot{\xi}_1^2(0) + \cdots + \dot{\xi}_N^2(0)} \,, \tag{74}$$

where the ξ's are the components of the tangent vector, i.e., of the perturbation of a reference trajectory, which obey the tangent dynamics equation (25). In the case of Eisenhart metric, each component of the Jacobi vector field J can be identified with the corresponding component of the

tangent vector ξ; moreover, ψ in Eq. (65) stands for *any* of the components of J, which obey the *same* effective equation. Thus, Eq. (74) becomes

$$\lambda = \lim_{t\to\infty} \frac{1}{2t} \log \frac{\psi^2(t) + \dot{\psi}^2(t)}{\psi^2(0) + \dot{\psi}^2(0)}\,, \tag{75}$$

where $\psi(t)$ is solution of Eq. (65). Eq. (75) is our estimate for the (largest) Lyapunov exponent.

As a stochastic differential equation, the solutions of Eq. (65) are properly defined after an averaging over the realizations of the stochastic process: referring to such an averaging as $\langle\bullet\rangle$, we rewrite Eq. (75) as

$$\lambda = \lim_{t\to\infty} \frac{1}{2t} \log \frac{\langle\psi^2(t)\rangle + \langle\dot{\psi}^2(t)\rangle}{\langle\psi^2(0)\rangle + \langle\dot{\psi}^2(0)\rangle}\,. \tag{76}$$

The evolution of $\langle\psi^2\rangle$, $\langle\dot{\psi}^2\rangle$ and $\langle\psi\dot{\psi}\rangle$, i.e., of the vector of the second moments of ψ, obeys the following equation, which can be derived by means of a technique, developed by Van Kampen and sketched in Appendix D:

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{pmatrix}\langle\psi^2\rangle\\ \langle\dot{\psi}^2\rangle\\ \langle\psi\dot{\psi}\rangle\end{pmatrix} = \begin{pmatrix} 0 & 0 & 2\\ \sigma_k^2\tau & 0 & -2k_0\\ -k_0 & 1 & 0\end{pmatrix}\begin{pmatrix}\langle\psi^2\rangle\\ \langle\dot{\psi}^2\rangle\\ \langle\psi\dot{\psi}\rangle\end{pmatrix}, \tag{77}$$

where k_0 and σ_k are the mean and the variance of $k(t)$, defined in Eqs. (66a) and (66b), respectively. Eq. (77) can be solved by diagonalizing the matrix on the r.h.s. of (77). The result for the evolution of $\langle\psi^2\rangle + \langle\dot{\psi}^2\rangle$ is

$$\langle\psi^2(t)\rangle + \langle\dot{\psi}^2(t)\rangle = (\langle\psi^2(0)\rangle + \langle\dot{\psi}^2(0)\rangle)\exp(\alpha t)\,, \tag{78}$$

where α is the only real eigenvalue of the matrix. According to Eq. (76), the Lyapunov exponent is given by $\lambda = \alpha/2$, so that, by computing explicitly α, one then finds the final expression

$$\lambda(k_0, \sigma_k, \tau) = \frac{1}{2}\left(\Lambda - \frac{4k_0}{3\Lambda}\right), \tag{79a}$$

$$\Lambda = \left(\sigma_k^2\tau + \sqrt{\left(\frac{4k_0}{3}\right)^3 + \sigma_k^4\tau^2}\right)^{1/3}. \tag{79b}$$

All the quantities k_0, σ_k and $\tau(k_0, \sigma_k)$ can be computed as *static* averages, as functions of the energy per degree of freedom, ε (see Eqs. (66a) and (66b)). Therefore – within the limits of validity of the assumptions made above – Eq. (79) provide an approximate analytic formula to compute the largest Lyapunov exponent independently of the numerical integration of the dynamics and of the tangent dynamics.

Let us remark that expanding Eqs. (79) in the limit $\sigma_k \ll k_0$ one finds that

$$\lambda \propto \sigma_k^2 \tag{80}$$

which shows how close the relation is between curvature fluctuations and dynamical instability.

4.2. Some applications

Let us now discuss briefly the results of the application of the geometric techniques described up to this point to some Hamiltonian models. In particular, we shall consider two cases: a chain of coupled non-linear oscillators (the so-called FPU β model, first introduced by Fermi et al. [68]) and a chain of coupled rotators (the 1-d XY model). The reason for the choice of these two particular models is that they allow fully analytic calculations and are well suited to show advantages and limitations of the theory. The geometric theory developed above has already been applied to many other cases, some of which will be addressed in Section 5. For other applications we refer to the literature: in particular, a model of a homopolymer chain has been studied in Ref. [32], a model of a three-dimensional Lennard–Jones crystal has been studied in Ref. [34], and a classical lattice gauge theory has been considered in Ref. [39]. The geometric approach has also been applied to a continuum field theory in one dimension [69]. Without entering into the details, we would like to single out a result which is shared by all the models considered until now. In all these models the functional dependence of the largest Lyapunov exponent on the energy per degree of freedom ε, in the low-ε limit, is numerically found to be

$$\lambda(\varepsilon) \propto \varepsilon^2 \,. \tag{81}$$

No explanation of this "universal" behaviour is yet at hand, and for some cases doubts about the validity of such a scaling with energy have been raised, because the numerical determination of Lyapunov exponents at low ε is difficult. However, the application of the geometric theory has provided a theoretical confirmation of this behaviour in all cases considered.

The systems we now consider are 1-d models with nearest-neighbour interactions whose Hamiltonians $\mathscr{H}$ have the standard form (1) with

$$V = \sum_{i=1}^{N} v(q_i - q_{i-1}) \,. \tag{82}$$

The interaction potentials are, respectively,

$$v(x) = \tfrac{1}{2}x^2 + \frac{u}{4}x^4 \quad (\text{FPU-}\beta \text{ model}) \,, \tag{83a}$$

or

$$v(x) = -J\cos x \quad (\text{1-d XY model}) \,. \tag{83b}$$

In Eq. (83a) we used u instead of the customary β in order to avoid confusion with the inverse temperature β. We assume $u > 0$.

The geometric quantities – in the framework of the Eisenhart metric – which are relevant in the quasi-isotropy approximation to describe the "effective" structure of the mechanical manifold, and which enter the geometric formula for the Lyapunov exponent, are the average and the root mean square (r.m.s.) fluctuations of the Ricci curvature of the mechanical manifold. They are defined as statistical averages computed in the *microcanonical* ensemble (see Eqs. (66)). First, we will show how these microcanonical quantities can be computed starting from the *canonical* partition function, which can be calculated exactly for an infinite chain, i.e., $N \to \infty$, for both models (83). Then, we will apply this procedure to each of the two models (83).

The average and fluctuations, within the microcanonical ensemble, of any observable function $f(q)$, can be computed as follows, in terms of the corresponding quantities in the canonical ensemble. The canonical configurational partition function $Z(\beta)$ is given by

$$Z(\beta) = \int \mathrm{d}q \mathrm{e}^{-\beta V(q)} \tag{84}$$

where $\mathrm{d}q = \prod_{i=1}^{N} \mathrm{d}q_i$. The canonical average $\langle f \rangle_{\mathrm{can}}$ of the observable f can be computed as

$$\langle f \rangle_{\mathrm{can}} = [Z(\beta)]^{-1} \int \mathrm{d}q f(q) \mathrm{e}^{-\beta V(q)} . \tag{85}$$

From this average, we can obtain the microcanonical average of f, $\langle f \rangle_{\mu}$, in the following (implicit) parametric form [70]:

$$\left.\begin{aligned} \langle f \rangle_{\mu}(\beta) &= \langle f \rangle_{\mathrm{can}}(\beta) \\ \varepsilon(\beta) &= \frac{1}{2\beta} - \frac{1}{N} \frac{\partial}{\partial \beta} [\log Z(\beta)] \end{aligned}\right\} \rightarrow \langle f \rangle_{\mu}(\varepsilon) . \tag{86}$$

Note that Eq. (86) is strictly valid only in the thermodynamic limit; at finite N, $\langle f \rangle_{\mu}(\beta) = \langle f \rangle_{\mathrm{can}}(\beta) + \mathcal{O}(1/N)$.

Contrary to the computation of $\langle f \rangle$, which is insensitive to the choice of the probability measure in the $N \rightarrow \infty$ limit, computing the fluctuations of f, i.e., of $\langle \delta^2 f \rangle = (1/N)\langle (f - \langle f \rangle)^2 \rangle$, by means of the canonical or microcanonical ensembles yields different results. The relationship between the canonical – i.e. computed with the Gibbsian weight $\mathrm{e}^{-\beta \mathscr{H}}$ – and the microcanonical fluctuations, is given by the Lebowitz–Percus–Verlet formula [70]

$$\langle \delta^2 f \rangle_{\mu}(\varepsilon) = \langle \delta^2 f \rangle_{\mathrm{can}}(\beta) - \frac{\beta^2}{c_V} \left[\frac{\partial \langle f \rangle_{\mathrm{can}}(\beta)}{\partial \beta} \right]^2 , \tag{87}$$

where

$$c_V = -\frac{\beta^2}{N} \frac{\partial \langle \mathscr{H} \rangle_{\mathrm{can}}}{\partial \beta} \tag{88}$$

is the specific heat at constant volume and $\beta = \beta(\varepsilon)$ is given in implicit form by the second equation in (86).

The average k_0 and the fluctuations σ_k of the Ricci curvature per degree of freedom are then obtained by replacing f with the explicit expression for Ricci curvature, which, according to the definition given in Eq. (64), is

$$K_R(q) = \sum_{i=1}^{N} \frac{\partial^2}{\partial q_i^2} v(q_i - q_{i-1}) \tag{89}$$

in Eqs. (86) and (87), respectively.

We now turn to the two applications mentioned above.

4.2.1. FPU β model

For the FPU β model the dynamical observable which corresponds to the Ricci curvature reads, according to Eq. (89),

$$K_R = 2N + 6u \sum_{i=1}^{N} (q_{i+1} - q_i)^2 \,. \tag{90}$$

Note that K_R is always positive and that this is also true for the sectional curvature along a physical geodesic. Computing the microcanonical average of K_R according to Eq. (86) we find that in the thermodynamic limit $k_0(\varepsilon)$ is implicitly given by (the details are reported in Ref. [31])

$$\left.\begin{aligned} \langle k_R \rangle_{\mathrm{can}}(\theta) &= 2 + \frac{3}{\theta}\frac{D_{-3/2}(\theta)}{D_{-1/2}(\theta)} \\ \varepsilon(\theta) &= \frac{1}{8\sigma}\left[\frac{3}{\theta^2} + \frac{1}{\theta}\frac{D_{-3/2}(\theta)}{D_{-1/2}(\theta)}\right] \end{aligned}\right\} \to k_0(\varepsilon)\,, \tag{91}$$

where the D_v are parabolic cylinder functions [63] and θ is a parameter proportional to β, so that $\theta \in [0, +\infty]$.

Let us now compute the fluctuations

$$\sigma_k^2(\varepsilon) = \frac{1}{N}\langle \delta^2 K_R \rangle_\mu(\varepsilon) = \frac{1}{N}\langle (K_R - \langle K_R \rangle)^2 \rangle_\mu \,. \tag{92}$$

According to Eq. (87), first the canonical fluctuation, $\langle \delta^2 k_R \rangle_{\mathrm{can}}(\beta) = (1/N)\langle (K_R - \langle K_R \rangle)^2 \rangle_{\mathrm{can}}(\beta)$, has to be computed and then a correction term must be added. For the canonical fluctuation we obtain [31]

$$\langle \delta^2 k_R \rangle_{\mathrm{can}}(\theta) = \frac{9}{\theta^2}\left\{2 - 2\theta\frac{D_{-3/2}(\theta)}{D_{-1/2}(\theta)} - \left[\frac{D_{-3/2}(\theta)}{D_{-1/2}(\theta)}\right]^2\right\} \tag{93}$$

and the final result for the fluctuations of the Ricci curvature is

$$\left.\begin{aligned} \langle \delta^2 k_R \rangle_\mu(\theta) &= \langle \delta^2 k_R \rangle_{\mathrm{can}}(\theta) - \frac{\beta^2}{c_V(\theta)}\left(\frac{\partial \langle k_R \rangle_{\mathrm{can}}(\theta)}{\partial \beta}\right)^2 \\ \varepsilon(\theta) &= \frac{1}{8\mu}\left[\frac{3}{\theta^2} + \frac{1}{\theta}\frac{D_{-3/2}(\theta)}{D_{-1/2}(\theta)}\right] \end{aligned}\right\} \to \sigma_k^2(\varepsilon)\,, \tag{94}$$

where $\langle \delta^2 k_R \rangle_{\mathrm{can}}(\theta)$ is given by (93), $\partial \langle k_R \rangle(\theta)/\partial\beta$ is given by

$$\frac{\partial \langle k_R \rangle(\theta)}{\partial \beta} = \frac{3}{8\mu\theta^3}\frac{\theta D_{-3/2}^2(\theta) + 2(\theta^2 - 1)D_{-1/2}(\theta)D_{-3/2}(\theta) - 2\theta D_{-1/2}^2(\theta)}{D_{-1/2}^2(\theta)} \tag{95}$$

and the specific heat per particle c_V is found to be [71,31]

$$\begin{aligned} c_V(\theta) = \frac{1}{16 D_{-1/2}^2(\theta)}\{&(12 + 2\theta^2)D_{-1/2}^2(\theta) + 2\theta D_{-1/2}(\theta)D_{-3/2}(\theta) \\ &- \theta^2 D_{-3/2}(\theta)[2\theta D_{-1/2}(\theta) + D_{-3/2}(\theta)]\} \,. \end{aligned} \tag{96}$$

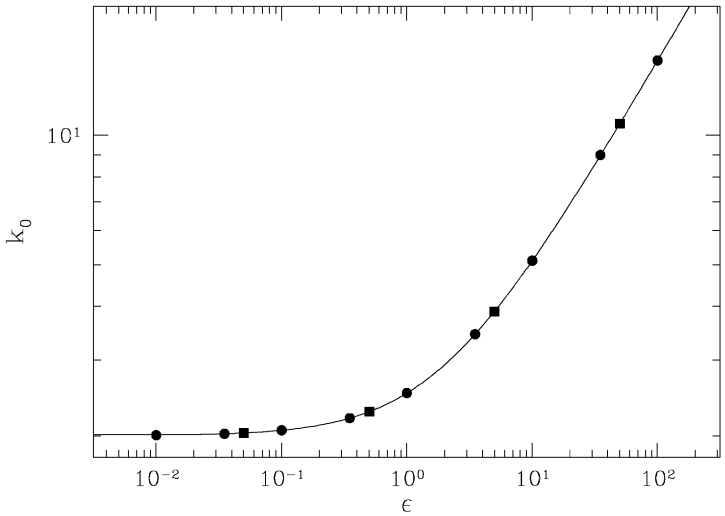

Fig. 1. Average Ricci curvature (Eisenhart metric) per degree of freedom, k_0, vs. energy density ε for the FPU-β model. The continuous line is the analytic computation according to Eq. (91); circles and squares are time averages obtained by numerical simulations with $N = 128$ and 512, respectively; $u = 0.1$. From Ref. [24].

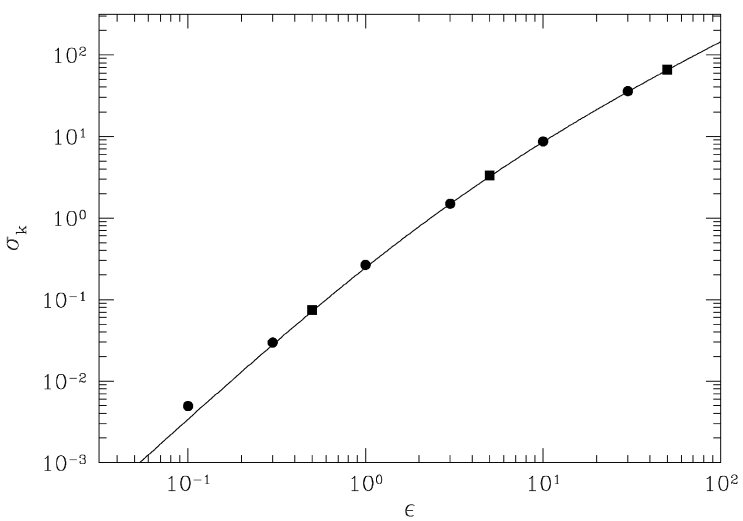

Fig. 2. Fluctuations of the Ricci curvature (Eisenhart metric), σ_k vs. energy density ε for the FPU-β model. Symbols and parameters as in Fig. 1; the continuous line now refers to Eq. (94). From Ref. [24].

The microcanonical averages and fluctuations computed in Eqs. (91) and (94) are compared in Figs. 1 and 2 with their corresponding time averages computed along numerically simulated trajectories of the FPU β-model with potential (83a) for $N = 128$ and 512 with $u = 0.1$. Though the microcanonical averages have to be computed in the thermodynamic limit, the agreement between time and ensemble averages is excellent already at $N = 128$.

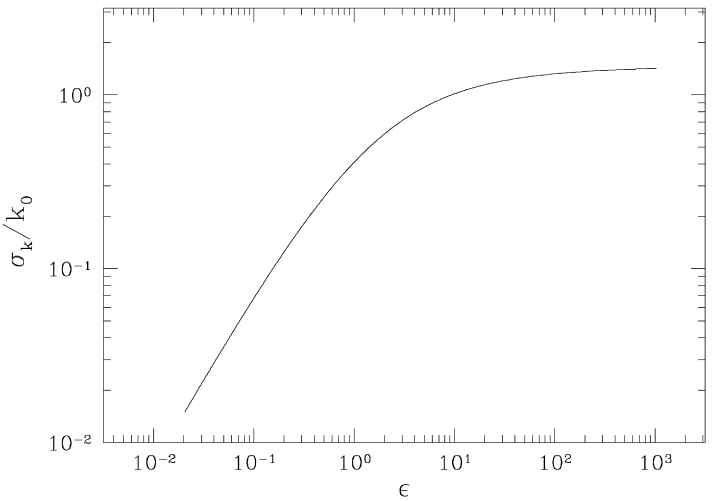

Fig. 3. Fluctuations of the Ricci curvature (Eisenhart metric) divided by the average curvature, σ_k/k_0, vs. energy density ε for the FPU-β model.

Before we comment on these results, we remark here that in many Hamiltonian dynamical systems different dynamical regimes can be found as the energy per degree of freedom ε is varied (see, for a review, Ref. [72] and references quoted therein). In particular, in the FPU-β model, a weakly chaotic regime is found for specific energies smaller than $\varepsilon_c \approx 0.1/u$ [73,74,24]. Although in the weakly chaotic regime the dynamics is chaotic (i.e., the Lyapunov exponent is positive, though small), mixing is very slow, as witnessed by the existence of a rather long memory of the initial conditions, i.e., of long relaxation times if the initial conditions are far from equilibrium. For ε larger than ε_c the dynamics is strongly chaotic and relaxations are fast. The precise origin of these phenomena is still to be understood. However, the geometric approach described here is able to provide a suggestive interpretation [24,33]. Let us consider Fig. 3, where the ratio of the fluctuations and the average curvature σ_k/k_0 is reported. As $\varepsilon \to 0$, $\sigma_k \ll k_0$, so that the manifold looks essentially like a constant curvature manifold with only small curvature fluctuations. This situation corresponds to the weakly chaotic dynamical regime. On the contrary, as ε is larger than ε_c, σ_k/k_0 tends to saturate towards a value of order unity, thus indicating that in the high-energy (strongly chaotic) regime the curvature fluctuations are of the same order of magnitude as the average curvature, so that the system no longer "feels" the isotropic (and integrable) limit. Hence, the geometric approach can give a hint for understanding, at least qualitatively, the origin of weak and strong chaos in the Fermi–Pasta–Ulam model.

The geometric theory also allows us to make a quantitative prediction for the Lyapunov exponent as a function of k_0 and σ_k via Eq. (79), which turns out to be extremely accurate. The analytic result is shown in Fig. 4 and is compared with numerical simulations made for different values of N, for the FPU-β case in a wide range of energy densities – more than six orders of magnitude [25,31]. The agreement between theory and simulations is remarkably good, which confirms the validity of the simplifying assumptions which we had to introduce on physical grounds to capture some of the essentials of the configuration space geometry.

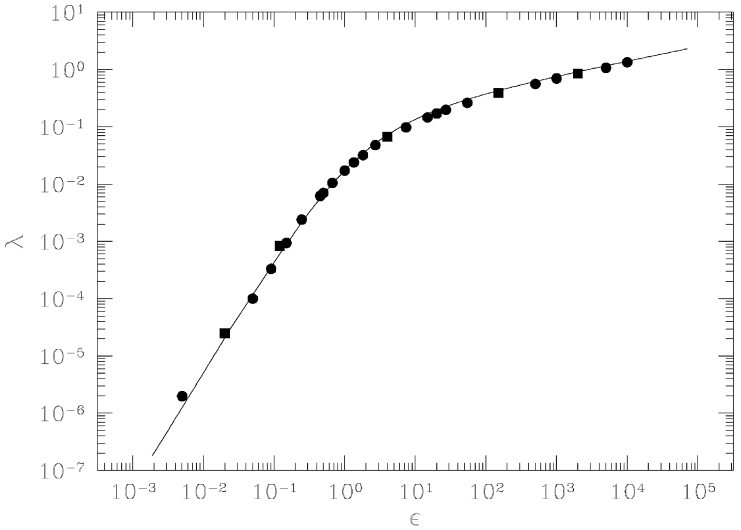

Fig. 4. Lyapunov exponent λ vs. energy density ε for the FPU-β model with $u = 0.1$. The continuous line is the theoretical computation according to Eq. (79), while the circles and squares are the results of numerical simulations with N, respectively, equal to 256 and 2000. From Ref. [31].

4.2.2. 1-d XY model

If the canonical coordinates q_i and p_i are given the meaning of angular coordinates and momenta, the 1-d XY model, whose potential energy is given in Eq. (83b), describes a linear chain of N rotators constrained to rotate on a plane and coupled by a nearest-neighbour interaction. This model can be formally obtained by restricting the classical Heisenberg model with O(2) symmetry to one spatial dimension. The potential energy of the O(2) Heisenberg model is $V = -J\sum_{\langle i,j\rangle} \boldsymbol{s}_i \cdot \boldsymbol{s}_j$, where the sum is extended only over nearest-neighbour pairs, J is the coupling constant and each $\boldsymbol{s}_i$ has unit modulus and rotates in the plane. To each "spin" $\boldsymbol{s}_i = (\cos q_i, \sin q_i)$, the velocity $\dot{\boldsymbol{s}}_i = (-\dot{q}_i \sin q_i, \dot{q}_i \cos q_i)$ is associated, so that $\mathscr{H} = \sum_{i=1}^{N} \frac{1}{2}\dot{\boldsymbol{s}}_i^2 - J\sum_{\langle i,j\rangle} \boldsymbol{s}_i \cdot \boldsymbol{s}_j$.

This Hamiltonian system has two integrable limits. In the small energy limit it represents a chain of harmonic oscillators, as can be seen by expanding the potential energy in a power series,

$$\mathscr{H}(p,q) \simeq \sum_{i=1}^{N} \left\{ \frac{p_i^2}{2} + J(q_{i+1} - q_i)^2 - 1 \right\}, \tag{97}$$

where $p_i = \dot{q}_i$, whereas in the high-energy limit a system of freely rotating objects is found, because the kinetic energy becomes much larger than the bounded potential energy.

The dynamics of this system has been extensively studied recently [71,75,28]. Numerical simulations and theoretical arguments independent of the geometric approach (see in particular Ref. [75]) have shown that also in this system there exist weakly and strongly chaotic dynamical regimes. It has been found that the dynamics is weakly chaotic in the low- and high-energy density regions, close to the two integrable limits. On the contrary, fully developed chaos is found in the intermediate-energy region.

According to Eq. (89), the expression of the Ricci curvature K_R, computed with the Eisenhart metric, is

$$K_R(q) = 2J \sum_{i=1}^{N} \cos(q_{i+1} - q_i). \tag{98}$$

We note that for this model a relation exists between the potential energy V and Ricci curvature K_R:

$$V(q) = JN - \frac{K_R(q)}{2}. \tag{99}$$

The average Ricci curvature can be again expressed by implicit formulae (see Ref. [31] for details)

$$\left.\begin{aligned} \langle k_R \rangle_\mu(\beta) &= 2J \frac{I_0(\beta J)}{I_1(\beta J)} \\ \varepsilon(\beta) &= \frac{1}{2\beta} + J\left(1 - \frac{I_1(\beta J)}{I_0(\beta J)}\right) \end{aligned}\right\} \to k_0(\varepsilon), \tag{100}$$

where the I_v's are modified Bessel functions of index v [63]. The fluctuations are given by the implicit equations

$$\left.\begin{aligned} \langle \delta^2 k_R \rangle(\beta) &= \frac{4J}{\beta} \frac{\beta J I_0^2(\beta J) - I_0(\beta J) I_1(\beta J) - \beta J I_1^2(\beta J)}{I_0^2(\beta J)[1 + 2(\beta J)^2] - 2\beta J I_1(\beta J) I_0(\beta J) - 2[\beta J I_1(\beta J)]^2} \\ \varepsilon(\beta) &= \frac{1}{2\beta} + J\left[1 - \frac{I_1(\beta J)}{I_0(\beta J)}\right] \end{aligned}\right\} \to \sigma_k^2(\varepsilon). \tag{101}$$

In Figs. 5 and 6 a comparison between analytical and numerical results is provided for the average Ricci curvature and its fluctuations. The agreement between ensemble and time averages is again very good.

Looking at Figs. 5 and 6 we realize that the low-energy weakly chaotic region has the same geometric properties as the corresponding region of the FPU model, as expected, since the two low-energy integrable limits are the same. On the contrary, in the high-energy weakly chaotic region the fluctuations are far from being small with respect to the average curvature. The average curvature $k_0(\varepsilon)$ vanishes as $\varepsilon \to \infty$. In this case the weakly chaotic dynamics seems related to the fact that the manifold $(M \times \boldsymbol{R}^2, g_E)$ looks almost flat along the physical geodesics. The bounds of the two weakly chaotic regions, as estimated in Ref. [75], coincide with the values of ε where the asymptotic behaviour of k (low-energy region) and σ_k (high-energy region) set in, respectively. Moreover, the case of the coupled rotators is very different from the FPU case, since the sectional curvature $K(s)$ along a geodesic can take negative values. The probability $P(\varepsilon)$ that $K(s) < 0$ can be analytically estimated in the following simple way. The explicit expression of the sectional curvature $K(\dot\gamma, \xi)$, relative to the plane spanned by the velocity vector $\dot\gamma = \mathrm{d}q/\mathrm{d}t$ and a generic vector $\xi \perp \dot\gamma$, is (see Eq. (A.43))

$$K(\dot\gamma, \xi) = R_{0i0k} \frac{\mathrm{d}q^0}{\mathrm{d}t} \frac{\xi^i}{\|\xi\|} \frac{\mathrm{d}q^0}{\mathrm{d}t} \frac{\xi^k}{\|\xi\|} \equiv \frac{\partial^2 V}{\partial q^i \partial q^k} \frac{\xi^i \xi^k}{\|\xi\|^2}, \tag{102}$$

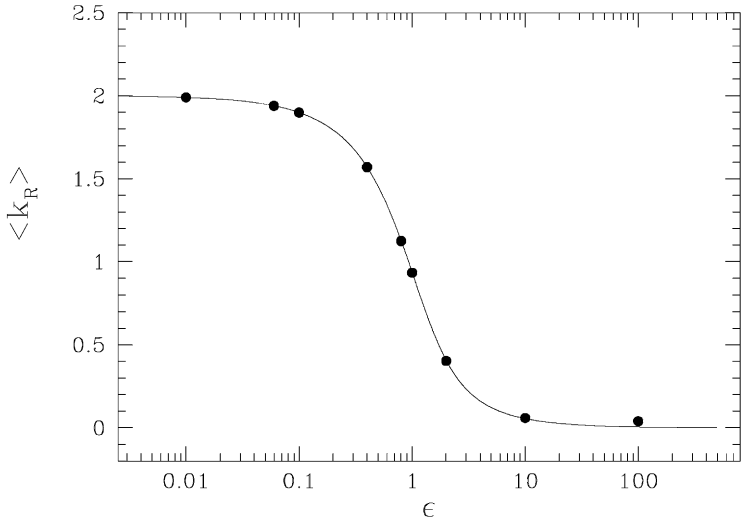

Fig. 5. Average Ricci curvature (Eisenhart metric) per degree of freedom k_0 vs. specific energy ε for the coupled rotators model. The continuous line is the result of an analytic computation according to Eq. (100); the full circles are time averages obtained by numerical simulations with $N = 150$; $J = 1$. From Ref. [31].

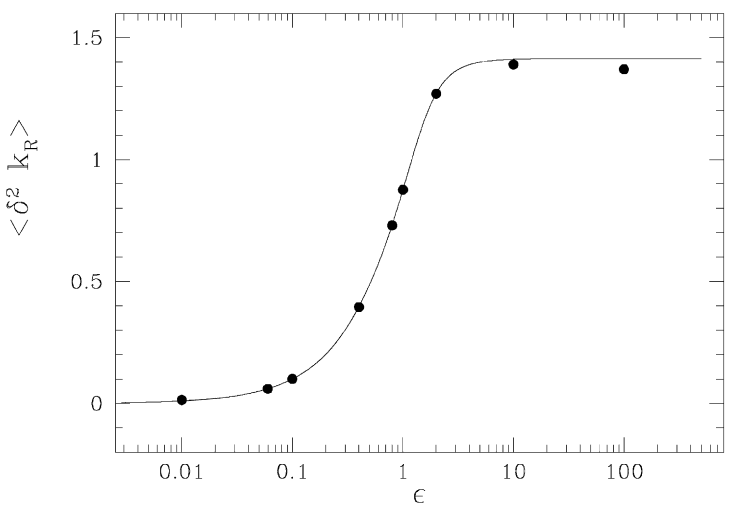

Fig. 6. Fluctuation of the Ricci curvature (Eisenhart metric) σ_k^2 vs. specific energy ε for the coupled rotators model. Symbols and parameters as in Fig. 5; the continuous line now refers to Eq. (101). From Ref. [31].

so that, computing $\partial^2 V/\partial q^i \partial q^k$ using the explicit form of $V(q)$ given in Eq. (83b), we get

$$K(\dot{\gamma}, \xi) = \frac{J}{\|\xi\|^2} \sum_{i=1}^{N} \cos(q_{i+1} - q_i)[\xi^{i+1} - \xi^i]^2 \tag{103}$$

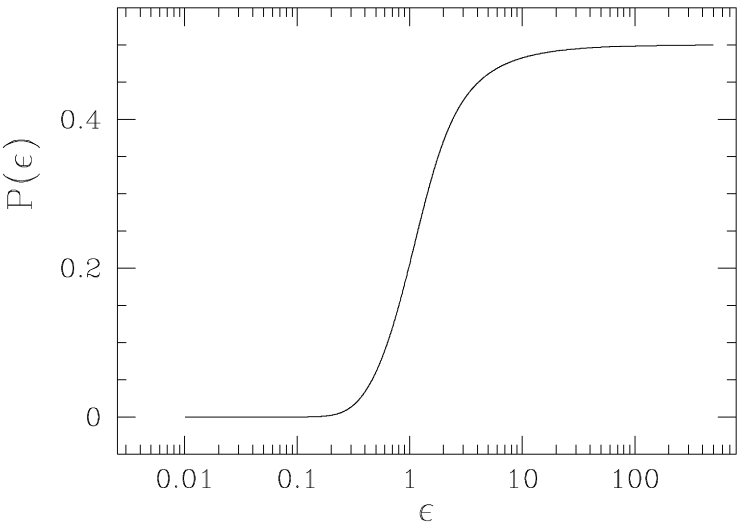

Fig. 7. Estimate of the probability $P(\varepsilon)$ of occurrence of negative sectional curvatures in the 1-d XY model according to Eq. (104); $J = 1$. From Ref. [31].

for the 1-d XY model. We realize, by simple inspection of Eq. (103), that the probability of finding $K < 0$ along a geodesic must be related to the probability of finding an angular difference larger than $\pi/2$ between two nearest-neighbouring rotators. From Eq. (103) we see that N orthogonal directions of the vector ξ exist such that the sectional curvatures – relative to the N planes spanned by these vectors together with $\dot{\gamma}$ – are just $\cos(q_{i+1} - q_i)$. These directions are defined by the unit vectors of components $(1, 0, \ldots, 0), (0, 1, 0, \ldots, 0), \ldots, (0, \ldots, 0, 1)$. Hence the probability $P(\varepsilon)$ of occurrence of a negative value of a cosine is used to estimate the probability of occurrence of negative sectional curvatures along the geodesics. This probability function, calculated using the Boltzmann weight, has the following simple expression [28,31]:

$$P(\varepsilon) = \frac{\int_{-\pi}^{\pi} \Theta(-\cos x) \mathrm{e}^{\beta J \cos x}\, \mathrm{d}x}{\int_{-\pi}^{\pi} \mathrm{e}^{\beta J \cos x}\, \mathrm{d}x} = \frac{\int_{\pi/2}^{3\pi/2} \mathrm{e}^{\beta J \cos x}\, \mathrm{d}x}{2\pi I_0(\beta J)}, \tag{104}$$

where $\Theta(x)$ is the Heavyside unit step function and I_0 the modified Bessel function of index 0. $P(\varepsilon)$ is plotted in Fig. 7. We see that in the strongly chaotic region such a probability starts to increase rapidly from a very small value, while it approaches an asymptotic value $P(\varepsilon) \simeq 1/2$ when the system enters its high-energy weakly chaotic region.

When the sectional curvatures are positive[9] chaos is produced by curvature fluctuations, hence we expect chaos to be weak as long as $\sigma_k/k_0 \ll 1$, and to become strong when $\sigma_k \approx k_0$. On the contrary, when $K(s)$ can assume both positive and negative values, the situation is much more

[9] The sectional curvature is always strictly positive in the FPU β model; in the 1-d XY model, in the low energy region, negative sectional curvatures can occur, but have a very small probability.

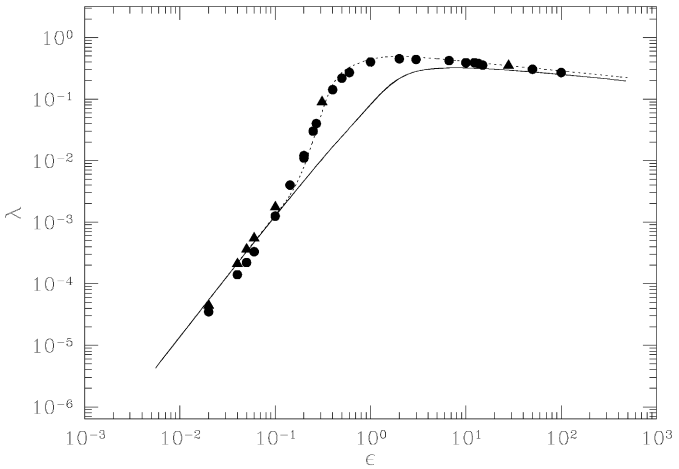

Fig. 8. Lyapunov exponent λ vs. energy density ε for the 1-d XY model with $J = 1$. The continuous line is the theoretical computation according to Eq. (79), while full circles, squares and triangles are the results of numerical simulations with N, respectively, equal to 150, 1000, and 1500. The dotted line is the theoretical result where the value of k entering Eq. (79) has been corrected according to Eq. (106) with $\alpha = 150$. From Ref. [31].

complicated, for there are now two different and independent sources of chaos: negative curvature which directly induces a divergence of nearby geodesics, and the bumpiness of the ambient manifold which induces such a divergence via parametric instability. The results for the coupled rotators model suggest that as long as the negative curvatures are "few" they do not dramatically change the picture, and may strengthen the parametrically generated chaos, while when their occurrence is equally likely as the occurrence of positive curvatures, the two mechanisms of chaos seem to inhibit each other and chaos becomes weak.[10]

Such a qualitative picture is consistent with the result of the geometric computation of λ for the coupled rotator model. The result of the application of Eq. (79) to this model is plotted in Fig. 8 (solid line). There is agreement between analytic and numeric values of the Lyapunov exponent only at low- and high-energy densities. Like in the FPU case, at low energy, in the quasi-harmonic limit, we find $\lambda(\varepsilon) \propto \varepsilon^2$. At high energy $\lambda(\varepsilon) \propto \varepsilon^{-1/6}$; here $\lambda(\varepsilon)$ is a decreasing function of ε because for $\varepsilon \to \infty$ the systems is integrable.

However, in the intermediate energy range our theoretical prediction underestimates the actual degree of chaos of the dynamics. This energy range coincides with the region of fully developed (strong) chaos. According to the above discussion the origin of the underestimation can be found in the fact that the role of the negative curvatures, which appears to strengthen chaos in this energy

[10] The fact that the two mechanisms, when comparable, can inhibit rather than strengthen each other can be considered a "proof" of the fact that their nature is intrinsically different. A similar situation is found also in some billiard systems, where there are two mechanisms which can produce chaos: (i) defocusing, due to positively curved boundaries, and (ii) divergence of the trajectories due to scatterings with negatively curved boundaries [76].

range, is not correctly taken into account. The sectional curvature $K(s)$, whose expression is given by Eq. (103), can take negative values with non-vanishing probability regardless of the value of ε, whereas, as long as $\varepsilon < J$, this possibility is lost in the replacement of K by the Ricci curvature, due to constraint (99), which implies that at each point of the manifold

$$k_R(\varepsilon) \geq 2(J - \varepsilon). \tag{105}$$

Thus, our approximation fails to account for the presence of negative sectional curvatures at values of ε smaller than J. In Eq. (103) the cosines have different and variable weights, $(\xi^{i+1} - \xi^i)^2$, which make it, in principle, possible to find somewhere along a geodesic a $K < 0$ even with only one negative cosine. This is not the case for k_R where all the cosines have the same weight.

Let us now show how the theoretical results can be improved. Our strategy is to modify the model for $K(s)$ in some *effective* way which takes into account the just mentioned difficulty of $k_R(s)$ to adequately model $K(s)$. This will be achieved by suitably "renormalizing" k_0 or σ_k to obtain an "improved" Gaussian process which can better model the behaviour of the sectional curvature. Since our "bare" Gaussian model underestimates negative sectional curvatures in the strongly chaotic region, the simplest way to renormalize the Gaussian process is to shift the peak of the distribution $\mathscr{P}(K_R)$ toward the negative axis to make the average smaller. This can easily be done by the following rescaling of the average curvature k_0:

$$k_0 = \langle k_R(\varepsilon) \rangle \to \frac{\langle k_R(\varepsilon) \rangle}{1 + \alpha P(\varepsilon)}. \tag{106}$$

This correction has no influence either when $P(\varepsilon) \simeq 0$ (below $\varepsilon \simeq 0.2$) or when $P(\varepsilon) \simeq 1/2$ (because in that case $\langle k_R(\varepsilon) \rangle \to 0$). The simple correction (106) makes use of the information we have obtained analytically, i.e., of the $P(\varepsilon)$ given in Eq. (104), and is sufficient to obtain an excellent agreement of the theoretical prediction with the numerical data over the whole range of energies, as shown in Fig. 8. The parameter α in (106) is a free parameter, and its value is determined so as to obtain the best agreement between numerical and theoretical data. The result shown in Fig. 8 (dotted line) is obtained with $\alpha = 150$, but also very different values of α, up to $\alpha \simeq 1000$, yield a good result, i.e., no particularly fine tuning of α is necessary to obtain a very good agreement between theory and numerical experiment.

4.3. Some remarks

Before moving to the second part of the paper, let us now comment about some points of the material presented in the first part of this Report. In particular, we would like to clarify the meaning of some of the approximations made and to draw the attention of the reader to some of the points which are still open. A better understanding of these points could lead, in our opinion, to a considerable improvement of the theory, which is still developing and can by no means be considered as a "closed" subject.

What has been presented in this section has a conceptual implication that goes far beyond the development of a method to analytically compute Lyapunov exponents. Rather, the strikingly good agreement between analytic and numeric Lyapunov exponents – obtained at the price of

a restriction of the domain of applicability[11] of the analytic expression worked out for λ – has three main implications:

(i) the *local* geometry of mechanical manifolds contains all the relevant information about (in)stability of Hamiltonian dynamics;
(ii) once a good model for the local source of instability of the dynamics is provided, then a statistical–mechanical-like treatment of the average degree of instability of the dynamics can be worked out, in the sense that we do not need a detailed knowledge of the dynamics but, by computing *global* geometric quantities, obtain a very good estimate of the average strength of chaos;
(iii) due to the variational formulation of newtonian dynamics, the Riemannian-geometric framework a priori seems – and actually seemed in the past (as we have pointed out in Section 2) – the natural framework to investigate the instability of Hamiltonian dynamics, although no evidence was available at all to confirm such an idea until the above mentioned development took place. It is now evident that the efforts to improve the theory by expanding its domain of applicability are worthwhile.

We must warn the reader though against "blind" applications of formula (79), i.e. without any idea about the fulfilment, by the Hamiltonian model under investigation, of the conditions under which it has been derived.

From a more technical point of view, one of the central results we have presented so far is the possibility of deriving, from the Jacobi equation, a *scalar* equation (Eq. (52)) describing the evolution of the Jacobi field J for a geodesic spread on a manifold. We would like to stress that such a result, though approximate, applies to a wide class of Hamiltonian systems. In fact, the only hypothesis needed to get such an equation is the quasi-isotropy hypothesis, i.e., the assumption that $R_{ijkl} \approx \mathscr{K}(s)(g_{ik}g_{jl} - g_{il}g_{jk})$. Loosely speaking, such an assumption means that, locally, the manifold can be regarded as isotropic, i.e., there is a neighbourhood of each point where the curvature can be considered constant. This does not imply at all that there are only small-scale fluctuations. There can be fluctuations of curvature on many scales, provided that they are finite and there is a cutoff at a certain point. The only case in which such an assumption will surely *not* hold is when there are curvature fluctuations over *all* scales. As will become clear in the following, this might happen when the manifold undergoes a topological change, and for "mechanical manifolds" this might happen at a phase transition.

Other approximations come into play when one actually wants to model $\mathscr{K}(s)$ along a geodesic with a stochastic process. It is true that replacing the sectional curvature by the Ricci curvature requires that the fluctuations are not only finite, but also small. Moreover, we use global averages to define the stochastic process, and here it is crucial that the fluctuations do not extend over too large scales. Thus Eq. (58) has a less general validity than Eq. (52). A way to improve the theory might be to try to replace the sectional curvature with some quantity related also to the gradient of the Ricci curvature, in order to make the replacement of sectional curvature less sensitive to the large scale variations of the Ricci curvature.

[11] In fact, the simplifying assumptions, made to work out an effective (in)stability equation from the full geodesic deviation equation, might be inadequate for some systems.

To get an explicit solution of Eq. (52), an even less general situation must be considered, through the following steps:

(i) using the Eisenhart metric;
(ii) considering standard systems where the kinetic energy does not depend on the q's;
(iii) estimating the characteristic correlation time τ of the curvature fluctuations.

As to item (iii), we have already remarked that our estimate given in Eq. (73) is by no means a consequence of any theoretical result, but only a reasonable estimate which could surely be improved.

As to (ii), the case of a more general kinetic energy matrix $a_{ij} \neq \delta_{ij}$, though not conceptually different, is indeed different in practice and the same final result is not expected to hold in that case.

Finally, item (i) should not reduce significantly the generality of the result. In fact, considering the Eisenhart metric only makes the calculations feasible, and in principle nothing should change, if one were able to solve Eq. (58) in the case of the Jacobi metric (see the discussion in Ref. [35]). However, Eisenhart and Jacobi metrics are *equivalent* for what concerns the computation of the average instability of the dynamics [35], but they might *not* be *equivalent* for other developments of the theory. This in view of the fact that (M_E, g_J) is a manifold which has better mathematical properties than $(M \times \boldsymbol{R}^2, g_E)$: (M_E, g_J) is a proper Riemannian manifold, it is compact, all of its geodesics are in one-to-one correspondence with mechanical trajectories, its scalar curvature does not identically vanish as is the case of $(M \times \boldsymbol{R}^2, g_E)$, it can be naturally lifted to the tangent bundle where the associated geodesic flow on the submanifolds of constant energy coincides with the phase space trajectories.

One could wonder whether the geometric approach is necessary at all to obtain our main results. Although in principle Eq. (52) could be obtained in many ways, using, e.g., assumptions about the distribution of the second derivatives of the potential, it is not clear how this should be implemented in practice, while here in the geometrical framework, a concrete scenario has been developed. In particular some physically reasonable guesses about the curvature and curvature correlation decay parameter τ can be made. Moreover, the good agreement with experiment gives some confidence that the approximations made, characterize indeed some of the main geometric properties of the highly complicated configuration space.

Let us finally add a comment on the application of the theory to the calculation of the Lyapunov exponent for the one-dimensional XY model. We have seen that although the predictions of the theory compare reasonably well with the numerical simulations, there is an intermediate energy range in which a correction must be added. As will become clear in the second part, the very first assumption (quasi-isotropy) should not be satisfied for this model, due to the presence of topology changes in the mechanical manifolds, in fact the difficulties of the theory begin just at the energy density which corresponds to the appearance of a large number of critical points of the potential energy (see next sections).

5. Geometry and phase transitions

In the previous sections we have shown how simple concepts belonging to classical differential geometry can be successfully used as tools to build a geometric theory of chaotic Hamiltonian

dynamics. Such a theory is able to describe the instability of the dynamics in classical systems consisting of a large number N of mutually interacting particles, by relating these properties to the average and the fluctuations of the curvature of the configuration space. Such a relation is made quantitative through Eq. (79), which provides an approximate analytical estimate of the largest Lyapunov exponent in terms of the above-mentioned geometric quantities, and which compares very well with the outcome of numerical simulations in a number of cases, two of which have been discussed in detail at the end of Section 4.

The macroscopic properties of large-N Hamiltonian systems can be understood by means of the traditional methods of statistical mechanics. One of the most striking phenomena that may happen in such systems is that when the external parameters (e.g., either the temperature or the energy) are varied until some critical value is reached, the macroscopic thermodynamical quantities may suddenly and even discontinuously change, so that, though the microscopic interactions are the same above and below the critical value of the parameters, its macroscopic properties may be completely different. Such phenomena are referred to as phase transitions. In statistical mechanics, phase transitions are explained as true mathematical singularities that occur in the thermodynamic functions in the limit $N \to \infty$, the so-called thermodynamic limit[12] [78]. Such singularities come from the fact that the equilibrium probability distribution in configuration space, which in the canonical ensemble is the Boltzmann weight

$$\varrho_{\mathrm{can}}(q_1, \ldots, q_N) = \frac{1}{Z} \exp[-\beta V(q_1, \ldots, q_N)], \tag{107}$$

where $\beta = 1/k_{\mathrm{B}}T$, V is the potential energy, and $Z = \int \mathrm{d}q\, \mathrm{e}^{-\beta V(q)}$ is the configurational partition function, can itself develop singularities in the thermodynamic limit.

The statistical–mechanical theory of phase transitions is one of the most elaborate and successful physical theories now at hand, and at least as continuous phase transitions are concerned, also quantitative predictions can be made, with the aid of renormalization-group techniques, which are in very good agreement with laboratory experiments and numerical simulations. We are not going to discuss this here, referring the reader to the vast literature on the subject [79–85].

However, the origin of the possibility of describing Hamiltonian systems via equilibrium statistical mechanics are the chaotic properties underlying the dynamics. In fact, though it is not possible to prove that generic Hamiltonian systems of the form (1) are ergodic and mixing, the fact that the trajectories are mostly chaotic (i.e., for the overwhelmingly majority of the trajectories positive Lyapunov exponents are found) means that such systems can be considered ergodic and mixing for all practical purposes.[13]

The observation that chaos is at the origin of the statistical behaviour of Hamiltonian systems and that chaotic dynamics can be described by means of the geometric methods described above, naturally leads to the following two questions:

[12] According to Uhlenbeck [77], the use of the thermodynamic limit as an explanation of the singularities of the partition function was suggested for the first time by Kramers in the 1938 Leiden conference on Statistical Mechanics.

[13] This has been recently extended to non-equilibrium statistical mechanics via the "chaotic hypothesis" [64].

1. Is there any specific behaviour of the Lyapunov exponent[14] when the system undergoes a phase transition?
2. What are the geometric properties of the configuration space manifold in the presence of a phase transition?

The aim of the present section is to discuss these two questions. We shall first give a phenomenological description which follows from numerical experiments, and then we shall concentrate on the particular case of the mean-field XY model where the geometrical quantities can be analytically calculated. From the discussion of these questions and from the (at least partial) answers that we find, we are lead to put forward a topological hypothesis about phase transitions, which will be discussed in Section 6.

5.1. Chaotic dynamics and phase transitions

In order to look for an answer to question 1 above, we now review the numerical results that have been obtained until now for various Hamiltonian dynamical systems which show a phase transition when considered as statistical–mechanical models for macroscopic systems in thermal equilibrium.

The first attempt to look for a chaotic-dynamic counterpart of an equilibrium phase transition is in the work by Butera and Caravati (BC) [86]. BC considered a two-dimensional XY model, i.e., a Hamiltonian dynamical system of the form (1) with the potential energy

$$V = 1 - \sum_{\langle i,j \rangle} \cos(\varphi_i - \varphi_j), \tag{108}$$

where the φ's are angles, i and j label the sites of a square lattice and the sum runs over all the nearest-neighbour sites. This model is the two-dimensional version of the one studied in Section 4.2.2. As the temperature is decreased, such a system undergoes a peculiar phase transition (referred to as the Berežinskij–Kosterlitz–Thouless, or BKT, transition) from a disordered phase to a quasi-ordered phase where, though no true long-range order is present, correlation functions decay as power laws, as occurs at a critical point [85]. Since there are no singularities in the finite-order derivatives of the free energy, the BKT transition is sometimes classified as an "infinite-order" phase transition. BC computed the Lyapunov exponent λ as a function of the temperature, and found that $\lambda(T)$ followed a rather smooth pattern; however, in a region around the transition, the dependence of λ on T changed from a steeply increasing function to a much less steep one.

BC's pioneering paper has been the only example of a study of this kind for a long period. However, very recently there has been a renewed interest in the study of the behaviour of Lyapunov exponents in systems undergoing phase transitions, and a number of papers appeared [36–38,87–97] The two-dimensional XY model has been reconsidered, together with the

[14] As in the first part of the paper, we consider only the largest Lyapunov exponent, which is referred to as just the Lyapunov exponent.

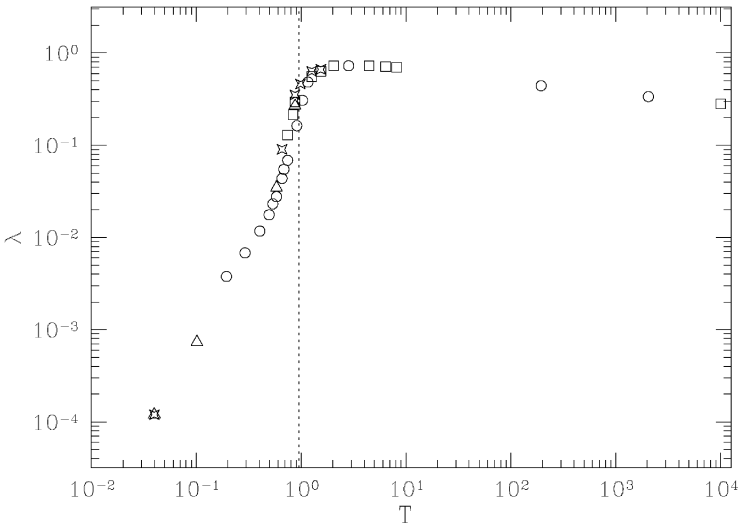

Fig. 9. Lyapunov exponent λ vs. the temperature T for the two-dimensional XY model: the circles refer to a 10×10, the squares to a 40×40, the triangles to a 50×50, and the stars to a 100×100 lattice, respectively. The critical temperature of the BKT transition is $T_C \simeq 0.95$ and is marked by a dotted vertical line. From Ref. [36].

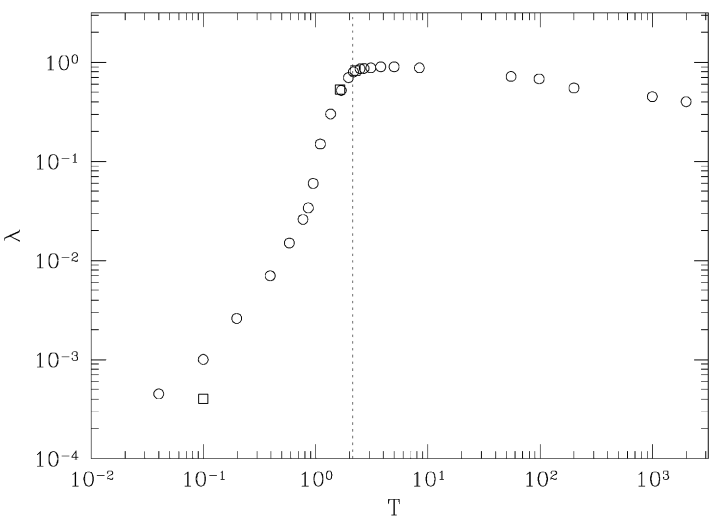

Fig. 10. Lyapunov exponent λ vs. the temperature T for the three-dimensional XY model, numerically computed on a $N = 10 \times 10 \times 10$ lattice (solid circles) and on a $N = 15 \times 15 \times 15$ lattice (solid squares), respectively. The critical temperature of the phase transition is $T_C \simeq 2.15$ and is marked by a dotted vertical line. From Ref. [36].

three-dimensional case, in Ref. [36]. We remark that in three spatial dimensions the XY model undergoes a standard continuous (second-order) phase transition accompanied by the breaking of the O(2) symmetry of the potential energy (108). The behaviour of the Lyapunov exponent λ as a function of the temperature T is shown in Figs. 9 and 10.

The behaviour found for the two-dimensional case confirms the BC results. The three-dimensional case shows a similar behaviour, but the change of the shape of the $\lambda(T)$ function near T_C is somehow sharper than in the previous case.

Dellago and Posch [91] considered an extended XY model, whose potential energy is

$$V = 2 - 2\sum_{\langle i,j\rangle} \cos\left(\frac{\varphi_i - \varphi_j}{2}\right)^{p^2}, \tag{109}$$

which includes the standard XY model (108) for $p^2 = 2$. On a two-dimensional lattice the transition, which is a continuous BKT transition for $p^2 = 2$, becomes a discontinuous transition as $p^2 \simeq 100$. The results for the Lyapunov exponent λ show that for any considered value of p^2 there is a change in the shape of $\lambda(T)$ close to the critical temperature.

One of the systems which have received considerable attention in this framework is the so-called lattice φ^4 model, i.e., a system with a Hamiltonian of the form (1) and a potential energy given by

$$V = \frac{J}{2}\sum_{\langle i,j\rangle} (\varphi_i - \varphi_j)^2 + \sum_i \left[-\frac{r^2}{2}\varphi_i^2 + \frac{u}{4!}\varphi_i^4 \right], \tag{110}$$

where the φ's are scalar variables, $\varphi_i \in [-\infty, +\infty]$, defined on the sites of a d-dimensional lattice, and r^2 and u are positive parameters. The lattice φ^4 model has a phase transition at a finite temperature provided that $d > 1$. The existence of such a transition, which belongs to the universality class of the d-dimensional Ising model, can be proved by renormalization-group arguments (see e.g. [80,98]). The cases $d = 2$ and 3 have been considered in Refs. [38,37]. Moreover, in Ref. [37] also some vector versions of this model have been considered, namely, systems with potential energy given by

$$V = \frac{J}{2}\sum_{\langle i,j\rangle}\sum_{\alpha} (\varphi_i^\alpha - \varphi_j^\alpha)^2 + \sum_i \left\{ -\frac{r^2}{2}\sum_\alpha (\varphi_i^\alpha)^2 + \frac{u}{4!}\left[\sum_\alpha (\varphi_i^\alpha)^2\right]^2 \right\}, \tag{111}$$

where α runs from 1 to n, labelling the components of the vectors $\varphi_i = (\varphi_i^1, \ldots, \varphi_i^n)$. The potential energy (111) is O(n)-invariant; in the case $n = 1$ we recover the scalar model (110). Figs. 11 and 12 show the behaviour of λ in the φ^4 model, in two and three dimensions, respectively.

Again we see that the Lyapunov exponent is sensitive to the presence of the transition, and that the shape of $\lambda(T)$ close to the transition is highly model-dependent. Moreover, such a shape can be significantly different within the same model as its parameters are varied. For instance, in the φ^4 model, λ can be either a monotonously increasing function of T or can display a maximum close to T_c, depending on the values of r^2 and u [38].

The Lyapunov exponents of systems undergoing phase transitions of the solid–liquid type have been recently determined numerically: Dellago and Posch (DP) considered, in two dimensions, a system of hard disks [89], a Lorentz-gas-like model and a Lennard–Jones fluid [90], and, in three dimensions, a system of hard spheres [92]. DP found that in all these systems the Lyapunov exponent is sensitive to the phase transition, and again the shape of λ is different for different models, the common feature being that λ attains a maximum close, if not at, the transition. Similar results have been obtained by Mehra and Ramaswamy [93]. Bonasera et al. [87] considered

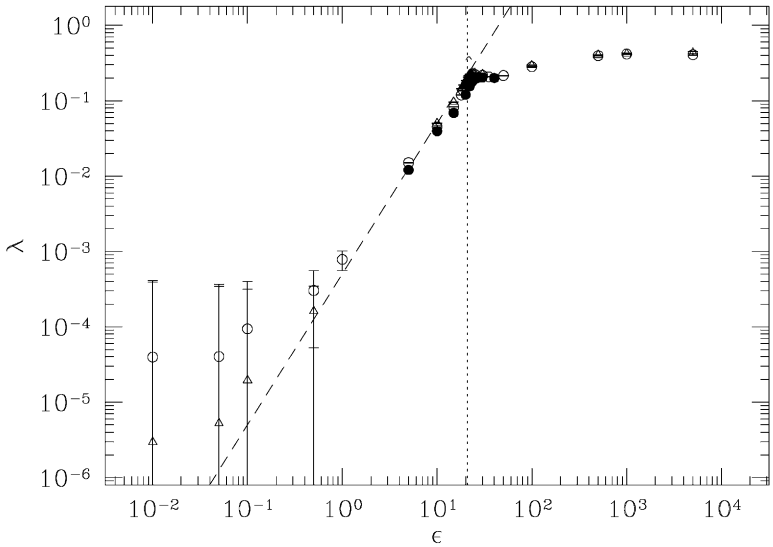

Fig. 11. Lyapunov exponent λ vs. the energy per particle ε, numerically computed for the two-dimensional O(1) φ^4 model, with $N = 100$ (solid circles), 400 (open circles), 900 (solid triangles), and 2500 (open triangles), respectively. The critical energy is marked by a vertical dotted line, and the dashed line is the power law ε^2. From Ref. [38].

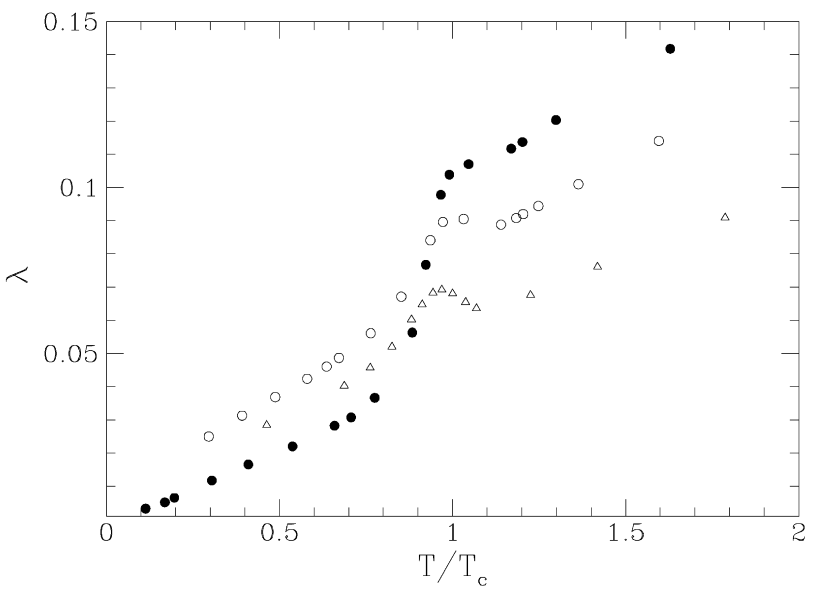

Fig. 12. Lyapunov exponent λ vs. the temperature T for the three-dimensional φ^4 model. Full circles correspond to the O(1) (scalar) case, open circles to the O(2) case, and open triangles to the O(4) case. From Ref. [37].

a classical model of an atomic cluster, whose particles interact via phenomenological pair potentials of the form

$$v(r) = a\,\mathrm{e}^{-(br/\sigma)} - c\left(\frac{\sigma}{r}\right)^6 \tag{112}$$

and of a nuclear cluster, with nucleons interacting via Yukawa pair potentials. Such systems undergo a so-called "multifragmentation" transition at a critical (model-dependent) temperature T_c. Bonasera et al. computed the Lyapunov exponents of these systems by means of numerical simulations at different temperatures. The resulting $\lambda(T)$ of both systems develops a sharp maximum close to T_c.

The numerical evidence that we have reviewed above clearly shows that the Lyapunov exponent of a Hamiltonian dynamical system is sensitive to the presence of a phase transition. However, the interpretation of the observed behaviour as it now stands is very difficult, because each model behaves differently and the behaviour of λ close to the transition does not apparently exhibit any universal feature: on the contrary, the shape of $\lambda(T)$ can depend on the values of the parameters of the model. Moreover, the qualitative behaviour of $\lambda(T)$ appears to be only weakly dependent on whether the transition is accompanied by a symmetry breaking or not, as in the case of the XY model: the shape of $\lambda(T)$ in two dimensions, where there is *not* any breaking of the O(2) symmetry of the potential energy below the BKT transition temperature, is not dramatically different from that of the three-dimensional case where the phase transition is accompanied by a symmetry breaking. In the latter case the "knee" of the $\lambda(T)$ curve is sharper, but it would be difficult to discriminate between the two cases only by looking at the $\lambda(T)$ curve. Therefore, though clearly sensitive to the presence of a phase transition, the Lyapunov exponent does not seem a "good" probing observable for the occurrence of a symmetry-breaking phase transition.

5.2. Curvature and phase transitions

In Section 4 we have seen that the origin of chaos in Hamiltonian mechanics can be understood from a geometrical point of view, and that the Lyapunov exponents are closely related to a geometric quantity, i.e., to the fluctuations of the Ricci curvature of the configuration space. Thus, it is natural to investigate whether such a geometric observable also has some peculiar behaviour close to the phase transition. As we shall see in the following, the fluctuations of the curvature do indeed have such a peculiar behaviour which, in turn, suggests a topological intepretation of the phase transition itself.

The Ricci curvature along a geodesic of the enlarged configuration space–time equipped with the Eisenhart metric, which we denoted by K_R in the previous sections, is given by the Laplacian of the potential energy – see Eq. (64). In the case of the XY model we obtain, as already shown in Section 4.2.2,

$$K_R = 2N - 2V = 2\sum_{\langle i,j\rangle} \cos(\varphi_i - \varphi_j)\,. \tag{113}$$

The root-mean-square fluctuation of K_R divided by the number of degrees of freedom N, i.e.,

$$\sigma_k = \left(\frac{1}{N}\langle K_R^2\rangle - \langle K_R\rangle^2\right)^{1/2}, \tag{114}$$

is plotted in Figs. 13 and 14 for the 2-d and 3-d cases, respectively.

In the case of the φ^4 model with O(n) symmetry, the Ricci curvature K_R is given by [37,38]

$$K_R = \sum_{\alpha=1}^{n}\sum_{i=1}^{N}\frac{\partial^2 V}{\partial(\varphi_i^\alpha)^2} = Nn(2Jd - r^2) + \lambda(n+2)\sum_{\alpha=1}^{n}\sum_{i=1}^{N}(\varphi_i^\alpha)^2\,. \tag{115}$$

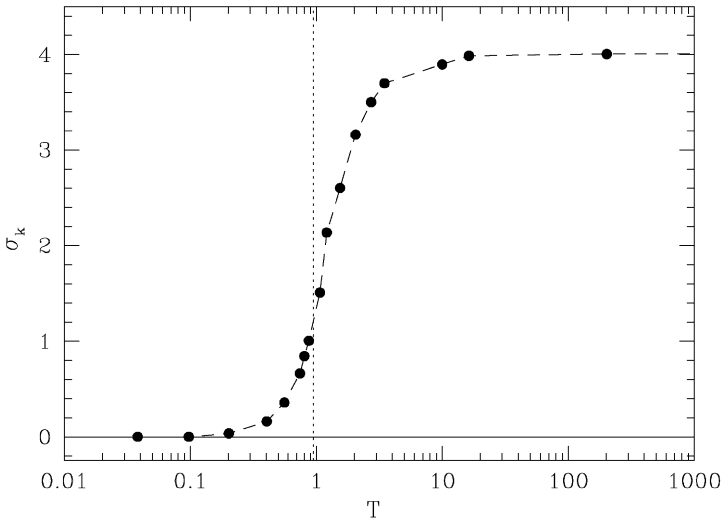

Fig. 13. Fluctuations of the Ricci curvature (Eisenhart metric), $\sigma_k(T)$, vs. the temperature T for the two-dimensional XY model. The solid circles are numerical values obtained for a 40×40 lattice; the dashed line is only a guide to the eye. The critical temperature of the BKT transition is $T_C \simeq 0.95$ and is marked by a dotted line. From Ref. [36].

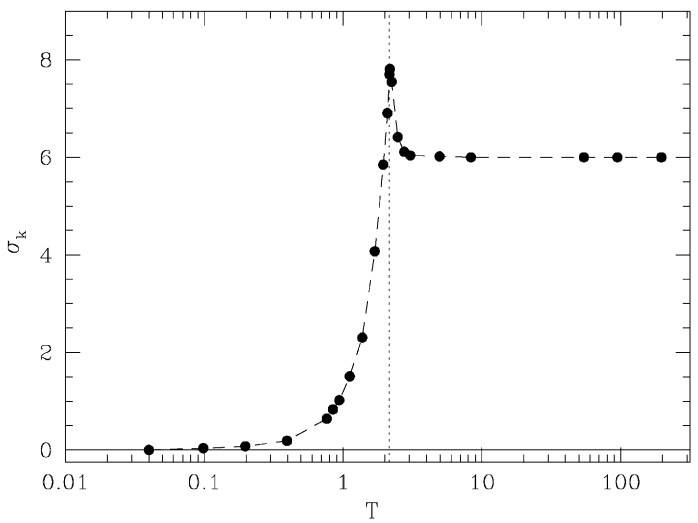

Fig. 14. As Fig. 13, for the three-dimensional XY model. Here $N = 10 \times 10 \times 10$, and the critical temperature of the phase transition is $T_c \simeq 2.15$. From Ref. [36].

The r.m.s. fluctuation of K_R, σ_k, is plotted against the energy per degree of freedom, ε, in the case of the two-dimensional O(1) φ^4 model in Fig. 15, and against the temperature T in the case of the two-dimensional O(2) φ^4 model in Fig. 16, and for the three-dimensional O(n) φ^4 models in Fig. 17.

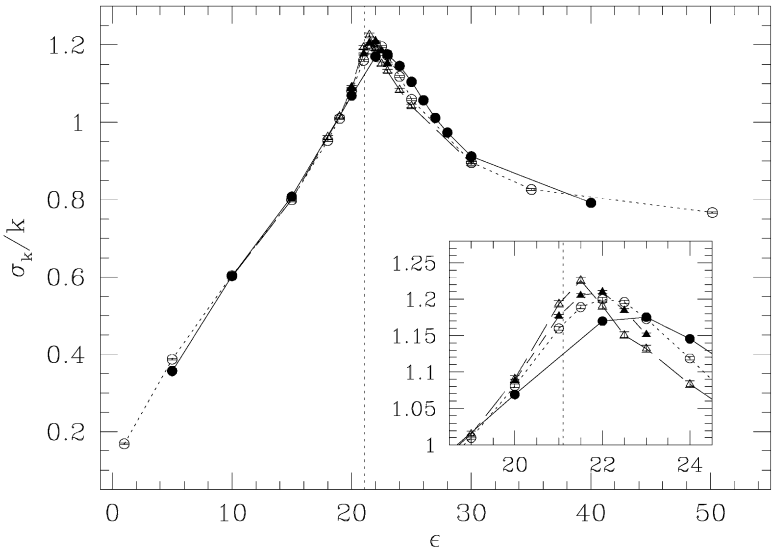

Fig. 15. Root mean square fluctuation of the Ricci curvature (Eisenhart metric) σ_k, divided by the average curvature k_0, numerically computed for the two-dimensional O(1) φ^4 model. The inset shows a magnification of the region close to the transition. Symbols as in Fig. 11. From Ref. [38].

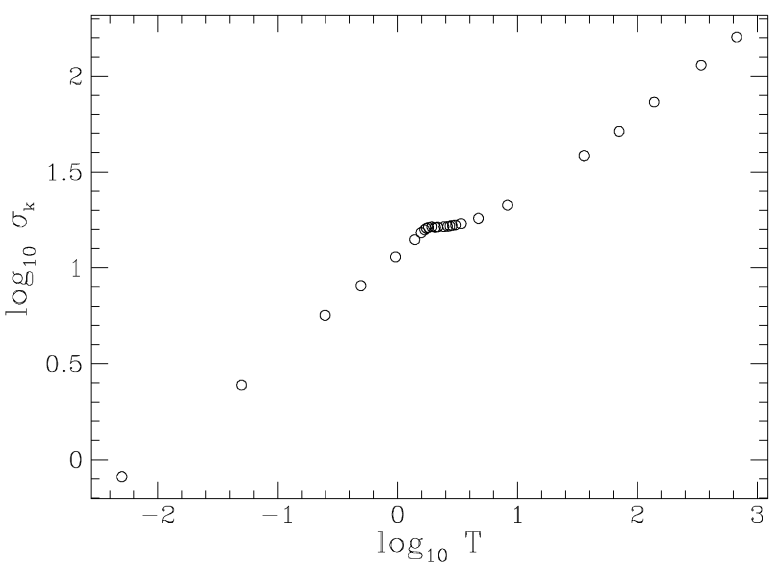

Fig. 16. Curvature fluctuations σ_k vs. the temperature T for the two-dimensional O(2) φ^4 model, numerically computed on a square lattice of 30×30 sites. The critical temperature T_c of the BKT transition is located at $T_c \simeq 1.5$. From Ref. [37].

Looking at Figs. 13–17, one can clearly see that when a symmetry-breaking phase transition occurs, a cusp-like ("singular") behaviour of the curvature fluctuations is found at the phase transition point (Figs. 14, 15 and 17), while, when only a BKT transition is present, no cusp-like

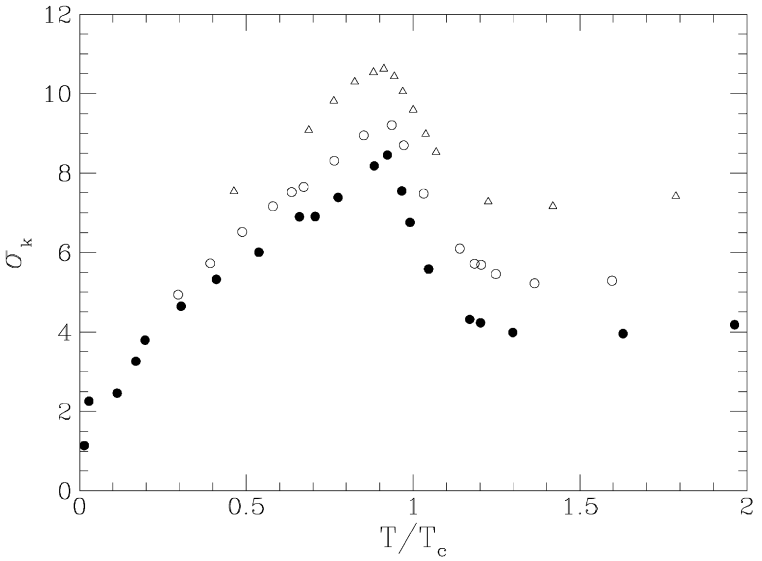

Fig. 17. Curvature fluctuations σ_k vs. the temperature T for the three-dimensional φ^4 model. Full circles correspond to the O(1) (scalar) case, open circles to the O(2) case, and open triangles to the O(4) case. From numerical simulations performed on an $8\times8\times8$ cubic lattice, reported in Ref. [37].

pattern is observed[15] (Figs. 13 and 16). We can summarize these results by saying that, in general, curvature fluctuations always show a cusp-like behaviour when a continuous symmetry-breaking phase transition is present, and, within numerical accuracy, the cusp occurs at the critical temperature. No counterexamples have yet been found to this general rule.

The fact that the Lyapunov exponent is sensitive to the phase transition can now be understood, in the light of the fact that, as shown in the first part of the present report, chaos can be described geometrically and, under rather general assumptions, the Lyapunov exponent is closely related to the fluctuations of the Ricci curvature (see Eq. (80)). However, contrary to the Lyapunov exponent, which, although sensitive to the phase transition, is not a good probing observable of the presence of a symmetry-breaking phase transition, the fluctuation of the Ricci curvature, σ_k, *is* a good probing observable, for, as plotted as a function of the temperature, it exhibits a clearly peculiar ("cuspy") pattern when a symmetry-breaking phase transition is present, and a rather smooth pattern otherwise. The difference among $\lambda(T)$ and $\sigma_k(T)$ as probing observables of the phase transition can be appreciated by comparing Figs. 12 and 17, where $\lambda(T)$ and $\sigma_k(T)$ are reported, respectively, for the O(n) φ^4 models. In Fig. 12, the $\lambda(T)$ curves for different values of n are *qualitatively* different, while in Fig. 17 the $\sigma_k(T)$ curves look strikingly similar, while being clearly different from the curve for the 2-d O(2) case (Fig. 16), where only a BKT transition is present. The same can be said in the case of the 2-d and the 3-d.

[15] Although the cusp-like behaviour is lost, indeed some change of behaviour is still visible in Figs. 13 and 16 close to the critical temperature, so that a BKT transition appears as "intermediate" between the absence of a phase transition and the presence of a symmetry-breaking phase transition.

5.2.1. Geometric estimate of the Lyapunov exponent

At this point, it is worthwile to point out that we can apply the geometric formula (79) for the Lyapunov exponent to estimate λ for all these models, since both k_0 and σ_k have been numerically computed. As shown in Refs. [36–38], one finds that in general, although the qualitative behaviour of the Lyapunov exponent is well reproduced, the quantitative agreement between the values of λ extracted from the numerical simulations and those obtained applying Eq. (79) is *not* good, in a neighbourhood of the phase transition.

However, this is to be expected, because among the assumptions under which formula (79) was derived there was the hypothesis that the fluctuations of the curvature should be not too large, and this is clearly not true close to a phase transition, as we have just shown.[16]

5.3. The mean-field XY model

The mean-field XY model [99] describes a system of N equally coupled planar classical rotators. It is defined by a Hamiltonian of class (1) where the potential energy is

$$V(\varphi) = \frac{J}{2N} \sum_{i,j=1}^{N} [1 - \cos(\varphi_i - \varphi_j)] - h \sum_{i=1}^{N} \cos \varphi_i \,. \tag{116}$$

Here $\varphi_i \in [0, 2\pi]$ is the rotation angle of the ith rotator and h is an external field. Defining at each site i a classical spin vector $\boldsymbol{s}_i = (\cos \varphi_i, \sin \varphi_i)$ the model describes a planar (XY) Heisenberg system with interactions of equal strength among all the spins. We consider only the ferromagnetic case $J > 0$; for the sake of simplicity, we set $J = 1$. The equilibrium statistical mechanics of this system is exactly described, in the thermodynamic limit, by mean-field theory [99]. In the limit $h \to 0$, the system has a continuous phase transition, with classical critical exponents, at $T_c = 1/2$, or $\varepsilon_c = 3/4$, where $\varepsilon = E/N$ is the energy per particle.

The Lyapunov exponent λ of this system is extremely sensitive to the phase transition. In fact, according to numerical simulations reported in Refs. [94,96,100,101], $\lambda(\varepsilon)$ is positive for $0 < \varepsilon < \varepsilon_c$, shows a sharp maximum immediately below the critical energy, and drops to zero at ε_c in the thermodynamic limit, where it remains zero in the whole region $\varepsilon > \varepsilon_c$, which corresponds to the thermodynamic disordered phase. In fact, in this phase the system is integrable, reducing to an assembly of uncoupled rotators. These results are valid in the thermodynamic limit $N \to \infty$ in the sense that they have been obtained by estimating the infinite N limit of finite N numerical simulations [94,96]: in the whole region $\varepsilon > \varepsilon_c$ the Lyapunov exponent, numerically computed for systems with different numbers of particles N, behaves as $\lambda \propto N^{-1/3}$, so that it extrapolates to zero at $N \to \infty$.

These results have received a theoretical confirmation in recent work by Firpo [95] based on the application of the geometric techniques described in the first part of the present paper. Firpo has computed analytically $\langle k_R \rangle$ and $\langle \delta^2 k_R \rangle$ in the thermodynamic limit for the mean-field XY model, showing that such quantities indeed have a singular behaviour at ε_c (see Fig. 18). Using these quantities and Eq. (79), Firpo has obtained the analytical estimate for $\lambda(\varepsilon)$ reported in Fig. 19; it is

[16] The results of formula (79) can be improved using procedures which are specific of the model under consideration and which we are not going to describe here (see Ref. [36] for the XY case and Ref. [37] for the φ^4 case, respectively).

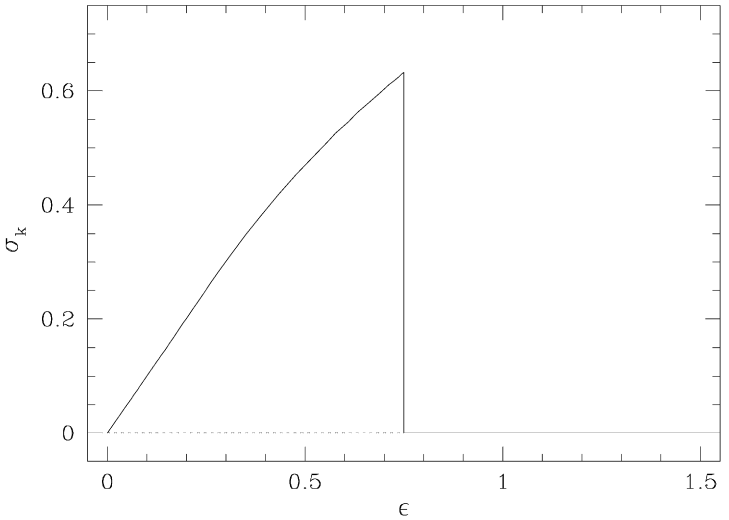

Fig. 18. Mean-field XY model: analytic expression for the microcanonical averages of the Ricci curvature (solid curve) and of its fluctuations (dot–dashed curve). From Ref. [95].

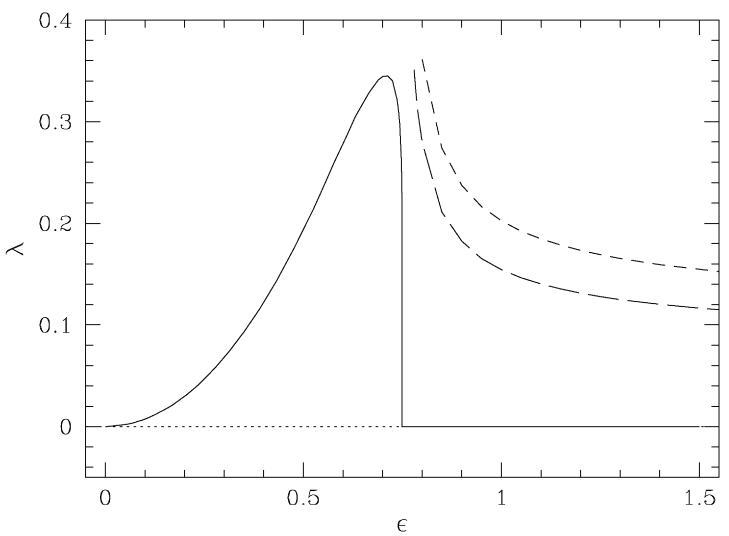

Fig. 19. Mean-field XY model: analytic expression for the Lyapunov exponent (solid curve). The curves above the transition are finite-N results for $N = 80$ and 200: here $\lambda \propto N^{-1/3}$. From Ref. [95].

remarkable that also the behaviour $\lambda \propto N^{-1/3}$ at $\varepsilon > \varepsilon_c$ has been extracted from this theoretical calculation. This result gives a theoretical confirmation to the qualitative behaviour of the Lyapunov exponent extrapolated from the numerical simulations. Moreover, Firpo's analytical results are in good quantitative agreement with numerical results reported in Refs. [96,100], also close to the phase transition and at variance with the cases of the nearest-neighbour XY and

φ^4 models considered earlier. A tentative explanation of why the application of the geometric formula (79) gives such good quantitative results in the present case can be that the mean-field character of the model prevents the curvature fluctuations from being too wild.

6. Phase transitions and topology

In the previous section we have reported results of numerical simulations for the fluctuations of observables of a geometric nature (e.g., configuration-space curvature fluctuations) related to the Riemannian geometrization of the dynamics in configuration space.[17] These quantities have been computed, using time averages, for many different models undergoing continuous phase transitions, namely φ^4 lattice models with discrete and continuous symmetries and XY models. In particular, when plotted as a function of either the temperature or the energy, the fluctuations of the curvature have an apparently singular behaviour at the transition point. Moreover, we have seen that the presence of a singularity in the statistical–mechanical fluctuations of the curvature at the transition point has been proved analytically for the mean-field XY model.

The aim of the present section is to try to understand on a deeper level the origin of this peculiar behaviour. In Section 6.1, we will show, using abstract geometric models, that a singular behaviour in the fluctuations of the curvature of a Riemannian manifold can be associated with a change in the *topology* of the manifold itself. By "topology change" we mean the following. Let us consider a surface $\mathscr{S}_\varepsilon$ which depends on a parameter ε in such a way that, upon varying the parameter, the surface is continuously deformed: as long as the different deformed surfaces can be mapped *smoothly* one onto another,[18] the topology does not change; on the contrary, the topology changes if there is a critical value of the parameter, say ε_c, such that the surface $\mathscr{S}_{\varepsilon>\varepsilon_c}$ cannot be mapped anymore smoothly onto $\mathscr{S}_{\varepsilon<\varepsilon_c}$.

The observation that a singularity in the curvature fluctuations of a Riemannian manifold, of the same type as those observed numerically at phase transitions, can be associated with a change in the topology of the manifold, leads us to conjecture that it is just this mechanism that could be at the basis of thermodynamic phase transitions. Such a conjecture was originally put forward in Ref. [36] as follows: a thermodynamic transition might be related to a change in the topology of the configuration space, and the observed singularities in the statistical–mechanical equilibrium measure and in the thermodynamic observables at the phase transition might be interpreted as a "shadow" of this major topological change that happens at a more basic level. We will refer to this conjecture as the *topological hypothesis* (TH).

The remaining part of the present section is devoted to a discussion of the TH and of its validity. In Section 6.2.1 we will report on a purely geometric, and thus still indirect, further indication that the topology of the configuration space might change at the phase transition, which has been obtained from numerical calculations for the φ^4 model on a two-dimensional lattice [102,103]. Then, in Section 6.2.3, we will describe a *direct* confirmation of the TH, i.e., we will show

[17] More precisely, we considered the enlarged configuration space-time, endowed with the Eisenhart metric.

[18] The different surfaces are then said to be diffeomorphic to each other (see Appendix B).

that a topological change in configuration space, which can be related with a phase transition, indeed occurs in the particular case of the mean-field XY model [104]. Finally, in Section 6.3 we will reformulate the TH in a more precise way, taking advantage of the previously discussed examples, and Section 6.4 will be devoted to a general discussion of the many points that are still open and on the future perspectives of the geometrical and topological approach to statistical mechanics.

6.1. *From geometry to topology*: *abstract geometric models*

Let us now describe how a singular behaviour of the curvature fluctuations of a manifold can be put in correspondence with a change in the topology of the manifold itself. For the sake of clarity, we shall first discuss a simple example concerning two-dimensional surfaces [36,37], and then we will generalize it to the case of N-dimensional hypersurfaces [102,103].

The simple geometric model we are going to describe concerns surfaces of revolution. A surface of revolution $\mathscr{S} \in \boldsymbol{R}^3$ is obtained by revolving the graph of a function f around one of the axes of a Cartesian plane, and can be defined, in parametric form, as follows [105]:

$$\mathscr{S}(u,v) \equiv (x(u,v), y(u,v), z(u,v)) = (a(u)\cos v, a(u)\sin v, b(u)), \tag{117}$$

where either $a(u) = f(u)$ and $b(u) = u$, if the graph of f is revolved around the vertical axis, or $a(u) = u$ and $b(u) = f(u)$, if the graph is revolved around the horizontal axis; in both cases, u and v are local coordinates on the surface $\mathscr{S}$: $v \in [0, 2\pi]$ and u belongs to the domain of definition of the function f.

Let us consider now, in particular, the two families of surfaces of revolution defined as

$$\mathscr{F}_\varepsilon = (f_\varepsilon(u)\cos v, f_\varepsilon(u)\sin v, u) \tag{118a}$$

and

$$\mathscr{G}_\varepsilon = (u\cos v, u\sin v, f_\varepsilon(u)), \tag{118b}$$

where

$$f_\varepsilon(u) = \pm\sqrt{\varepsilon + u^2 - u^4}, \qquad \varepsilon \in [\varepsilon_{\min}, +\infty), \tag{119}$$

and $\varepsilon_{\min} = -\frac{1}{4}$. Some cases are shown in Fig. 20.

There exists for both families of surfaces a critical value of ε, $\varepsilon_c = 0$, corresponding to a change in the *topology* of the surfaces: the manifolds $\mathscr{F}_\varepsilon$ are diffeomorphic to a torus $\boldsymbol{T}^2$ for $\varepsilon < 0$ and to a sphere $\boldsymbol{S}^2$ for $\varepsilon > 0$; the manifolds $\mathscr{G}_\varepsilon$ are diffeomorphic to *two* spheres for $\varepsilon < 0$ and to one sphere for $\varepsilon > 0$. The Euler–Poincaré characteristic (see Eq. (B.10)) is $\chi(\mathscr{F}_\varepsilon) = 0$ if $\varepsilon < 0$, and $\chi(\mathscr{F}_\varepsilon) = 2$ otherwise, while $\chi(\mathscr{G}_\varepsilon)$ is 4 or 2 for ε negative or positive, respectively.

We now turn to the definition and the calculation of the curvature fluctuations on these surfaces. Let M belong to one of the two families; its Gaussian curvature K is [105]

$$K = \frac{a'(a''b' - b'a'')}{a(b'^2 + a'^2)^2}, \tag{120}$$

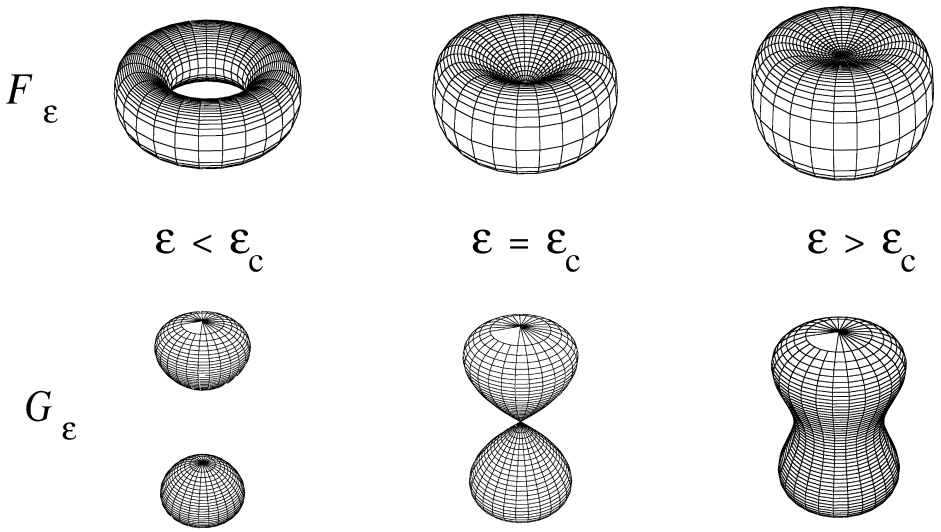

Fig. 20. Some representatives of the two families of surfaces $\mathscr{F}_\varepsilon$ and $\mathscr{G}_\varepsilon$ defined in Eqs. (118a) and (118b), respectively. Each family is divided into two subfamilies by the critical surface corresponding to $\varepsilon_c = 0$ (middle members in the picture). Members of the same subfamily are diffeomorphic, whereas the two subfamilies are not diffeomorphic to each other. From Ref. [37].

where $a(u)$ and $b(u)$ are the coefficients of Eq. (117), and primes denotes differentiation, with respect to u. The fluctuations of K can be then defined as

$$\sigma_K^2 = \langle K^2 \rangle - \langle K \rangle^2 = A^{-1} \int_M K^2 \, \mathrm{d}S - \left(A^{-1} \int_M K \, \mathrm{d}S \right)^2 , \tag{121}$$

where A is the area of M and $\mathrm{d}S$ is the invariant surface element. Both families of surfaces exhibit a singular behaviour in σ_K as $\varepsilon \to \varepsilon_c$, as shown in Fig. 21, in spite of their different curvature properties on the average.[19]

We are now going to show that the result we have just obtained for two-dimensional surfaces has a much more general validity: a *generic* topology change in an n-dimensional manifold is accompanied by a singularity in its curvature fluctuations [102]. In order to do that, we have to make use of some concepts belonging to Morse theory, which will also be used in Section 6.2.3 below; the main concepts of Morse theory are sketched in Appendix B, where also references to the literature are given.

We consider then a hypersurface of $\boldsymbol{R}^N$ which is the u-level set of a function f defined in $\boldsymbol{R}^N$, i.e., a submanifold of $\boldsymbol{R}^N$ of dimension $n = N - 1$ defined by the equation

$$f(x_1, \ldots, x_N) = u , \tag{122}$$

[19] For instance, $\langle K \rangle(\varepsilon) = 0$ for $\mathscr{F}_\varepsilon$ as $\varepsilon < 0$, while for $\mathscr{G}_\varepsilon$ the same average curvature is positive and diverges as $\varepsilon \to 0$.

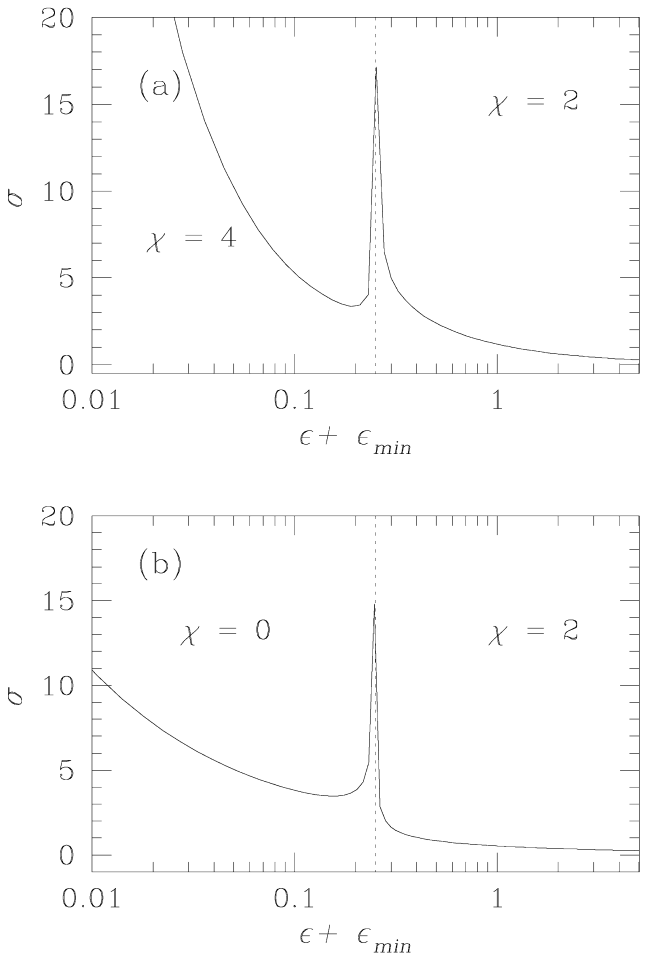

Fig. 21. The fluctuation σ_K of the Gaussian curvature of the surfaces $\mathscr{F}_\varepsilon$ and $\mathscr{G}_\varepsilon$ is plotted vs. ε. σ is defined in Eq. (121), ε is shifted by $\varepsilon_{\min} = 0.25$ for reasons of clarity of presentation. (a) refers to $\mathscr{G}_\varepsilon$ and (b) refers to $\mathscr{F}_\varepsilon$. The cusps appear at $\varepsilon = 0$ where the topological transition takes place for both $\mathscr{F}_\varepsilon$ and $\mathscr{G}_\varepsilon$. From Ref. [37].

such a hypersurface can then be referred to as $f^{-1}(u)$. Let us now assume[20] that f is a *Morse function*, i.e., such that its critical points (i.e., the points of $\boldsymbol{R}^N$ where the differential df vanishes) are isolated. One of the most important results of Morse theory is that the topology of the hypersurfaces $f^{-1}(u)$ can change *only* crossing a critical level $f^{-1}(u_c)$, i.e., a level set containing at least one critical point of f. This means that a generic change in the topology of the hypersurfaces can be associated with critical points of f. Now, the hypersurfaces $f^{-1}(u)$ can be given a Riemannian metric in a standard way [106], and it is possible to analyze the behaviour of the curvature fluctuations in

[20] This is not a strong assumption: in fact, it can be shown that Morse functions are generic (see Appendix B).

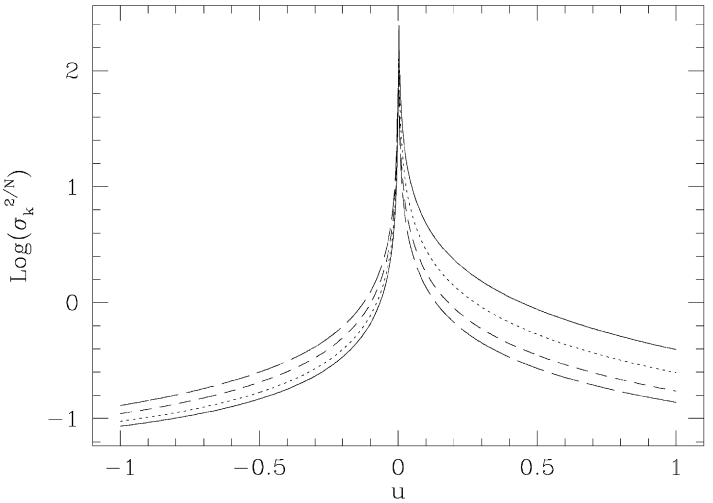

Fig. 22. Fluctuations of the Gauss curvature of a hypersurface $f^{-1}(u)$ of $\boldsymbol{R}^N$ vs. u close to a critical point. $\sigma_K^{2/N}$ is reported because it has the same dimensions of the scalar curvature. Here $\dim(f^{-1}(u)) = 100$, and the Morse indexes are: $k = 1, 15, 33, 48$, represented by solid, dotted, dashed, long-dashed lines, respectively.

a neighborhood of a critical point. Let us assume, for the sake of simplicity, that this critical point is located at $x_0 = 0$ and belongs to the level $u_c = 0$. Any Morse function can be parametrized, in the neighborhood of a x_0, by means of the so-called Morse chart, i.e., a system of local coordinates $\{y_i\}$ such that $f(y) = f(x_0) - \sum_{i=1}^{k} y_i^2 + \sum_{i=k+1}^{N} y_i^2$ (k is the Morse index of the critical point). Then standard formulae for the Gauss curvature K of hypersurfaces of $\boldsymbol{R}^N$ [106] can be used to compute explicitly the fluctuations of the curvature, σ_K, of the level set $f^{-1}(u)$. Numerical results for the curvature fluctuations are reported in Fig. 22 and show that also at high dimension σ_K^2 develops a sharp, singular peak as the critical surface is approached (for computational details, see Ref. [103]).

6.2. Topology changes in configuration space and phase transitions

As we have discussed in Section 5, the curvature fluctuations of the configuration space exhibit cusp-like patterns in presence of a second-order phase transition. A truly cuspy pattern, i.e., an analytic discontinuity, is mathematically proven in the case of mean-field XY model. In Section 6.1, we have shown that singular patterns in the fluctuations of the curvature of a Riemannian manifold can be seen as consequences of the presence of a topology change. Hence, we are led to the topological hypothesis (TH), i.e., to conjecture that at least continuous, symmetry-breaking phase transitions are associated with topology changes in the configuration space of the system.

However, an important question arises, in that the fluctuations of the curvature considered in Section 5 have been obtained as time averages, computed along the dynamical trajectories of the Hamiltonian systems under investigation (or as statistical averages computed analytically, as in the case of the mean-field XY model). Now, time averages of geometric observables are usually found to be in excellent agreement with ensemble averages [24,31,36–38] so that one could argue that the

above-mentioned singular-like patterns of the fluctuations of geometric observables are simply the precursors of truly singular patterns due to the fact that the measures of all the statistical ensembles tend to become singular in the limit $N \to \infty$ when a phase transition is present. In other words, geometric observables, like any other "honest" observable, already at finite N would feel the eventually singular character of the statistical measures, i.e., of the probability distribution functions of the statistical–mechanical ensembles. If this were the correct explanation, we could not attribute the cusp-like patterns of the curvature fluctuations to any special geometric features of configuration space, and the cusp-like patterns observed in the numerical simulations could not be considered as (indirect) confirmations of the TH.

In order to elucidate this important point, three different paths have been followed: (i) *purely geometric* information about certain submanifolds of configuration space has been worked out *independently* of the statistical measures in the case of the two-dimensional φ^4 model, and the results lend *indirect* support to the TH [102]; (ii) a *direct* numerical confirmation of the TH has been given in [107] by means of the computation of a topologic invariant, the Euler characteristic, in the case of a 2d lattice φ^4 model; (iii) a *direct* analytic confirmation of the TH has been found in the particular case of the mean-field XY model [104]. We report on items (i), (ii) and (iii) in Sections 6.2.1, 6.2.2 and 6.2.3, respectively.

6.2.1. *Indirect numerical investigations of the topology of configuration space*

In order to separate the singular effects due to the singular character of statistical measures at a phase transition from the singular effects due to some topological transition in configuration space, the first natural step is to consider again σ_K^2 as an observable, and to integrate it on suitable submanifolds of configuration space by means of a geometric measure, i.e. by means of a measure which has nothing to do with statistical ensemble measures.

Consider as ambient space the N-dimensional configuration space M of a Hamiltonian system with N degrees of freedom, which, when $N \to \infty$, undergoes a phase transition at a certain finite temperature T_c (or critical energy per degree of freedom ε_c), and let $\mathscr{V} = V(\varphi)$ be its potential energy per degree of freedom.

Then the relevant geometrical objects are the submanifolds of M defined by

$$M_u = \mathscr{V}^{-1}(-\infty, u] = \{\varphi \in M : \mathscr{V}(\varphi) \le u\}, \tag{123}$$

i.e., each M_u is the set $\{\varphi_i\}_{i=1}^N$ such that the potential energy per particle does not exceed a given value u. As u is increased from $-\infty$ to $+\infty$, this family covers successively the whole manifold M. All the submanifolds M_u can be given a Riemannian metric g whose choice is largely arbitrary. On all these manifolds (M_u, g) there is a standard invariant volume measure

$$\mathrm{d}\eta = \sqrt{\det(g)}\,\mathrm{d}q^1 \ldots \mathrm{d}q^N, \tag{124}$$

which has nothing to do with statistical measures. Let us finally define the hypersurfaces Σ_u as the u-level sets of $\mathscr{V}$, i.e.,

$$\Sigma_u = \mathscr{V}^{-1}(u), \tag{125}$$

which are nothing but the boundaries of the submanifolds M_u.

According to the discussion reported in Section 6.1, an indirect way to study the presence of topology changes in the family $\{(M_u, g)\}$ is to look at the behaviour of the fluctuations of the Gaussian curvature, σ^2_K, defined as

$$\sigma^2_K = \langle K^2_G \rangle_{\Sigma_u} - \langle K_G \rangle^2_{\Sigma_u} , \tag{126}$$

where $\langle \cdot \rangle$ stands for integration over the surface Σ_u, as a function of u. The presence of cusp-like singularities of σ^2_K for some critical value of u, u_c, would eventually signal the presence of a topology change of the family $\{(M_u, g)\}$ at u_c [102]. Such an indirect geometric probing of the presence of critical points seems an expedient way to probe the possible topology changes of the manifolds (M_u, g). In fact, the properties of the manifolds M_u are closely related to those of the hypersurfaces $\{\Sigma_u\}_{u \le u_c}$, as can be inferred from the equation

$$\int_{M_u} f \mathrm{d}\eta = \int_0^u \mathrm{d}v \int_{\Sigma_v} f|_{\Sigma_v} \, \mathrm{d}\omega / \|\nabla V\| , \tag{127}$$

where $\mathrm{d}\omega$ is the induced measure[21] on Σ_u and f a generic function [108]. From Morse theory (see Appendix B) we know that the surface Σ_{u_c} defined by $\mathscr{V} = u_c$ is a degenerate quadric, so that in its vicinity some of the principal curvatures [106] of the surfaces $\Sigma_{u \simeq u_c}$ tend to diverge.[22] Such a divergence is generically detected by any function of the principal curvatures and thus, for practical computational reasons, instead of the Gauss curvature (which is the product of all the principal curvatures) we shall consider the total second variation of the *scalar* curvature $\mathscr{R}$ (i.e., the sum of all the possible products of two principal curvatures) of the manifolds (M_u, g), according to the definition

$$\sigma^2_{\mathscr{R}}(u) = [Vol(M_u)]^{-1} \int_{M_u} \mathrm{d}\eta \left[\mathscr{R} - [Vol(M_u)]^{-1} \int_{M_u} \mathrm{d}\eta \mathscr{R} \right]^2 \tag{128}$$

with $\mathscr{R} = g^{kj} R^l_{kij}$, where R^l_{kij} are the components of the Riemann curvature tensor [see Eq. (A.39)] and $Vol(M_u) = \int_{M_u} \mathrm{d}\eta$. The subsets M_u of configuration space are given the structure of Riemannian manifolds (M_u, g) by endowing all of them with the *same* metric tensor g. However, the choice of the metric g is arbitrary in view of probing possible effects of the topology on the geometry of these manifolds.

In Ref. [102] the configuration spaces of a φ^4 model, defined on a one-dimensional lattice (linear chain) and on a two-dimensional square lattice have been considered. We recall that in the 2-d case this system undergoes a phase transition at a finite temperature, while in the 1-d case no phase transition is present. Three different types of metrics have been considered, i.e.,

(i) $g^{(1)}_{\mu\nu} = [A - V(\varphi)]\delta_{\mu\nu}$, i.e., a conformal deformation (see Appendix A.3) of the Euclidean flat metric $\delta_{\mu\nu}$, where $A > 0$ is an arbitrary constant chosen large enough to be sure that in the relevant interval of values of u the determinant of the metric is always positive;

[21] If a surface is parametrically defined through the equations $x^i = x^i(z^1, \ldots, z^k)$, $i = 1, \ldots, N$ then the metric g_{ij} induced on the surface is given by $g_{ij}(z^1, \ldots, z^k) = \sum_{n=1}^N (\partial x^n / \partial z^i)(\partial x^n / \partial z^j)$.

[22] The principal curvatures are the inverse of the curvature radii measured, at any given point of a surface, in suitable directions. At a Morse critical point some of these curvature radii vanish.

(ii) $g^{(2)}_{\mu\nu}$ and $g^{(3)}_{\mu\nu}$ are generic metrics (no longer conformal deformations of the flat metric) defined by

$$(g^{(k)}_{\mu\nu}) = \begin{pmatrix} f^{(k)} & 0 & 1 \\ 0 & \boldsymbol{I} & 0 \\ 1 & 0 & 1 \end{pmatrix}, \quad k = 2, 3, \tag{129}$$

where $\boldsymbol{I}$ is the $(N-2)$-dimensional identity matrix, $g^{(2)}$ is obtained by setting $f^{(2)} = (1/N)\sum_{\alpha\in\boldsymbol{Z}^d}\varphi^4_\alpha + A$, and $g^{(3)}$ by setting $f^{(3)} = (1/N)\sum_{\alpha\in\boldsymbol{Z}^d}\varphi^6_\alpha + A$, with $A > 0$, and α labels the N lattice sites of a linear chain ($d = 1$) or of a square lattice ($d = 2$, $N = n \times n$).

These choices are completely arbitrary, however, and only if metrics of very simple form are chosen, both analytical and numerical computations are feasible also for rather large values of N. Thus, the first metric has been chosen diagonal, and the other two metrics concentrate in only one matrix element all the non-trivial geometric information. Moreover, the first metric still contains a reference to the physical potential, whereas the other two define metric structures that are completely independent of the physical potential and only contain monomials of powers sufficiently high that they do not vanish after two successive derivatives have been taken (needed to compute curvatures). The topology of the subsets of points M_u and Σ_u of $\boldsymbol{R}^N$ is already determined (though well concealed) by the definitions of Eqs. (123) and (125); the task is to "capture" some information about their topology through a mathematical object or structure, defined on these sets of points, which is capable of mirroring the variations of topology through the u-pattern of an analytic function. This idea follows the philosophy of standard mathematical theories of differential topology, for example, within Morse theory, the information about topology is extracted through the critical points of any function – defined on a given manifold – fulfilling some conditions (necessary to be a good Morse function), or, within cohomology theory [109], topology is probed through vector spaces of differential forms (the de Rham's cohomology vector spaces) "attached" to a given manifold. Provided that good mathematical quantities are chosen as topology-variation detectors, arbitrary Riemannian metric structures could work as well.

For the above-defined metrics $g^{(k)}$, $k = 1, 2, 3$, simple algebra leads from the definition of the scalar curvature (see Appendix A) to the following explicit expressions:

$$\mathscr{R}^{(1)} = (N-1)\left[\frac{\Delta V}{(A-V)^2} - \frac{\|\nabla V\|^2}{(A-V)^3}\left(\frac{N}{4} - \frac{3}{2}\right)\right], \tag{130}$$

$$\mathscr{R}^{(k)} = \frac{1}{(f^{(k)}-1)}\left[\frac{\|\tilde{\nabla} f^{(k)}\|^2}{2(f^{(k)}-1)} - \tilde{\Delta} f^{(k)}\right], \quad k = 2, 3 \tag{131}$$

where ∇ and Δ are the Euclidean gradient and Laplacian, respectively, and $\tilde{\nabla}$ and $\tilde{\Delta}$ lack the derivative $\partial/\partial\varphi_\alpha$ with $\alpha = 1$ in the $d = 1$ case, and lack the derivative $\partial/\partial\varphi_\alpha$ with $\alpha = (1, 1)$ in the $d = 2$ case.

The numerical computation of the geometric integrals in Eq. (128) is worked out by means of a Monte Carlo algorithm [33,103] to sample the geometric measure $d\eta$ by means of an "importance sampling" algorithm [110] suitably modified. In Figs. 23 and 24 and $\sigma_{\mathscr{R}}(u)$ are given for the one and two-dimensional cases obtained for two different lattice sizes with $g^{(1)}$ (Fig. 23), and at

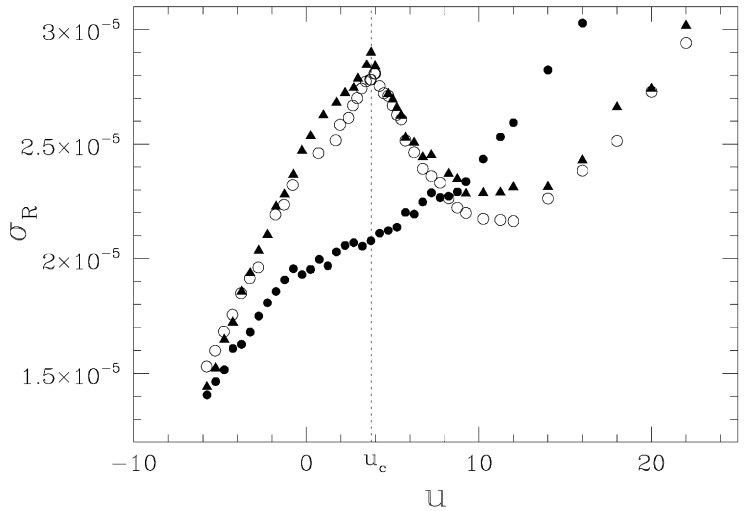

Fig. 23. Variance of the scalar curvature of M_u vs. u computed with the metric $g^{(1)}$. Full circles correspond to the 1d-φ^4 model with $N = 400$. Open circles refer to the 2d-φ^4 model with $N = 20 \times 20$ lattice sites, and full triangles refer to 40×40 lattice sites (whose values are rescaled for graphic reasons).

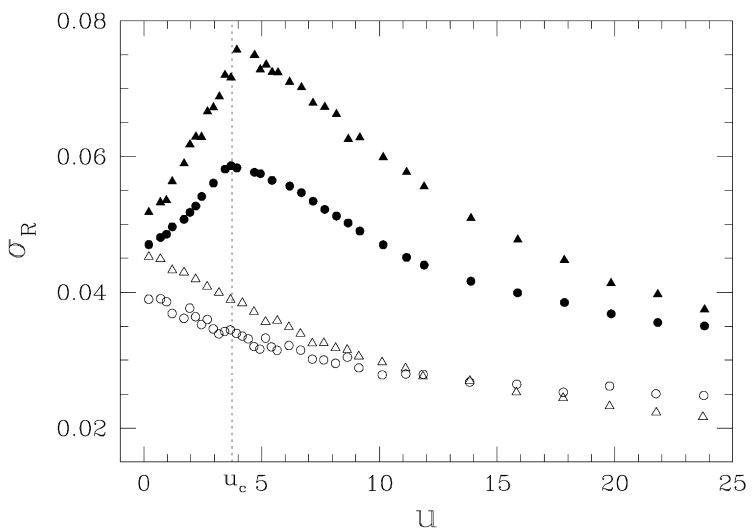

Fig. 24. $\sigma_{\mathscr{R}}(u)$ of M_u vs. u computed for the φ^4 model with: metric $g^{(2)}$ in 1d, $N = 400$ (open triangles); metric $g^{(2)}$ in 2d, $N = 20 \times 20$ (full triangles); metric $g^{(3)}$ in 1d, $N = 400$ (open circles); metric $g^{(3)}$ in 2d, $N = 20 \times 20$ (full circles).

given lattice size with $g^{(2,3)}$ (Fig. 23). Peaks of $\sigma^2_{\mathscr{R}}(u)$ appear at a certain value v_c, of v in the two-dimensional case, whereas only smooth patterns are found in the one-dimensional case, where no phase transition is present.

According to the discussion above, these peaks can be considered as indirect evidence of the presence of a topology transition in the manifolds M_u at $u = u_c$ in the case of the two-dimensional φ^4 model. It is, in particular, the persistence of cusp-like patterns of $\sigma_{\mathscr{R}}(u)$ independently of the

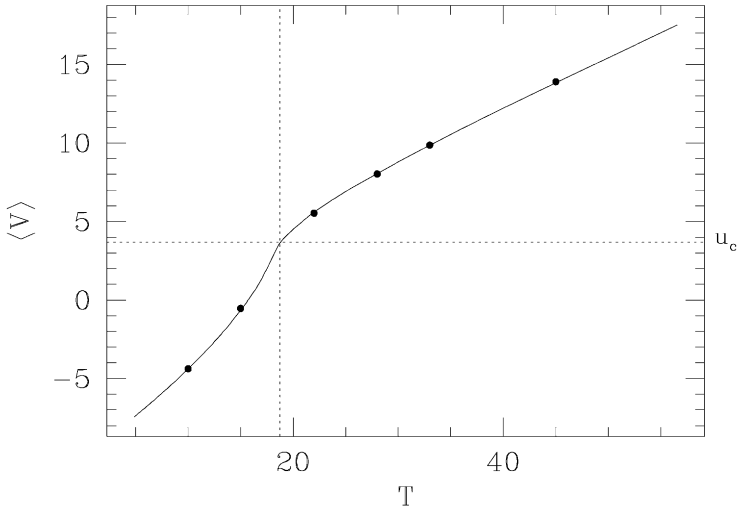

Fig. 25. Average potential energy vs. temperature for the 2-d lattice φ^4 model with O(1) symmetry. Lattice size $N = 20 \times 20$. The solid line is made out of 200 points obtained as time averages. Full circles represent Monte Carlo estimates of canonical ensemble averages. The dotted lines locate the phase transition.

metric chosen that lends credit to the idea that this actually reflects a topological transition. Now, we want to argue that the topological transition occurring at u_c is related to a *thermodynamic phase transition* which occurs in the φ^4 model. In order to do that, in Ref. [102] the average potential energy per particle

$$u(T) = \langle \mathscr{V} \rangle \tag{132}$$

has been numerically computed, as a function of T, by means of both Monte Carlo averaging with the canonical configurational measure, and Hamiltonian dynamics. In the latter case the temperature T is given by the average kinetic energy per degree of freedom, and u is obtained as time average. Fig. 25 shows a perfect agreement between time and ensemble averages. The phase transition point is well visible at $u = u_c \simeq 3.75$. Looking at Figs. 23 and 24, we realize that, within the numerical accuracy, the critical value of the potential energy per particle u_c where the topological change occurs equals the statistical–mechanical average value of the potential energy at the phase transition.

At this point the doubt, formulated at the beginning of this section, about the possible non-geometrical origin of the "singular", cusp-like patterns of $\sigma_{\mathscr{H}}(u)$ has been dissipated. These results have been found *independently* of statistical mechanical measures and of their singular character in presence of a phase transition. These results are also *independent* – at least to the limited extent of the three metric tensors reported above – of the geometric structure given to the family $\{M_u\}$. Thus, they seem most likely to have their origin at a deeper level than the geometric one, i.e. at the topologic level. Hence the observed phenomenology strongly hints that some *major* change in the topology of the configuration-space-submanifolds $\{M_u\}$ occurs when a second-order phase transition takes place.

6.2.2. Direct numerical investigation of the topology of configuration space

Though still based on numerical computations for a special model, a *direct* evidence of the tight relation between topology and phase transitions has been obtained by computing the u-dependence of a topologic invariant of the leaves Σ_u in the foliation of configuration space into a family of equipotential surfaces.

In order to directly probe if and how the topology change – in the sense of a breaking of *diffeomorphicity* of the surfaces Σ_u – is actually the counterpart of a phase transition, a *diffeomorphism invariant* has to be computed. This is a very challenging task because of the high dimensionality of the manifolds involved. Moreover, any algorithm of a combinatorial type (like those implied by simplicial decompositions, i.e. high-dimensional analogs of tesselations with triangles that are used, for example, in numerical quantum gravity for low-dimensional manifolds) is here hopeless. Only through a link between analytic and topologic mathematical objects can one hope to work out some direct information about topology. One such a link is provided by the Gauss–Bonnet–Hopf theorem that relates the Euler characteristic (see Appendix B) $\chi(\Sigma_u)$ with the total Gauss–Kronecker curvature of the manifold, i.e. [105]

$$\chi(\Sigma_u) = \gamma \int_{\Sigma_u} K_G \, \mathrm{d}\sigma \tag{133}$$

which is valid in general for even dimensional hypersurfaces of Euclidean spaces $\boldsymbol{R}^N$ [here $\dim(\Sigma_u) = n \equiv N - 1$], and where $\gamma = 2/Vol(\boldsymbol{S}^n_1)$ is twice the inverse of the volume of an n-dimensional sphere of unit radius; K_G is the Gauss–Kronecker curvature of the manifold; $\mathrm{d}\sigma = \sqrt{det(g)}\mathrm{d}x^1\,\mathrm{d}x^2 \cdots \mathrm{d}x^n$ is the invariant volume measure of Σ_u and g is the Riemannian metric induced from $\boldsymbol{R}^N$. The Gauss–Kronecker curvature at a given point of a hypersurface is the product of the eigenvalues of its so-called shape operator; these eigenvalues are the principal curvatures of the hypersurface at the given point. The shape operator is constructed through the directional derivatives of the unit normal vector to the hypersurface at the given point computed in the n directions of the basis vectors of the plane tangent to the surface at the same point [106].

The numerical application of the Gauss–Bonnet–Hopf theorem has been worked out for a lattice φ^4 model in one and two spatial dimensions [103,115,107]. The main result is shown in Fig. 26 where $\chi(\Sigma_v)$ is reported vs. v. In the 1d case (open circles) a "smooth" pattern of $\chi(v)$ is found, whereas in the 2d case a cusp-like shaped $\chi(v)$ shows up with the singular point corresponding to the phase transition point marked by the vertical dotted line. The parameters are those of the preceding section, therefore the phase transition point is at $v_c/N \simeq 3.75$.

These results have two important consequences: (i) the non-constant value of $\chi(v)$ in the 1d case clearly shows that topology changes are present even in the absence of phase transitions; (ii) an *abrupt change* in the rate of variation of topology with v seems the hallmark of a phase transition.[23] Thus, we have direct numerical evidence about the actual implication of topology in the appearance of a phase transition. At the same time we have evidence of a non-simple one-to-one correspondence between topology changes and phase transitions. This is in full agreement with

[23] The fact that an "elementary" topology change is not sufficient to entail a phase transition and that it is necessary an abrupt change in the way of changing of topology has been discussed in Ref. [50].

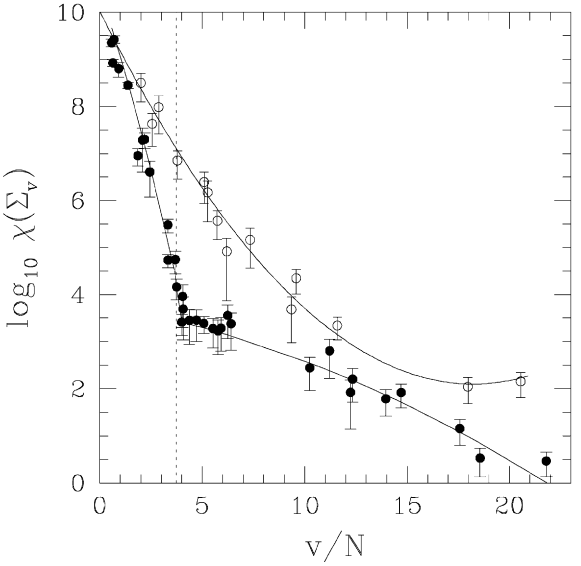

Fig. 26. Euler characteristic $\chi(\Sigma_v)$ for 1-d and 2-d φ^4 lattice models. Open circles: 1-d case, $N = 49$; full circles: 2-d case, $N = 7 \times 7$. The vertical dotted line, computed separately for larger N, accurately locates the phase transition and the parameters are the same as in Figs. 23, 24 and 25.

what is discussed in the next section about a different kind of potential which can be analytically investigated.

6.2.3. Topological origin of the phase transition in the mean-field XY model

Until now we have not yet given any *direct* analytic evidence of the validity of the TH. Let us now consider again the mean-field XY model (116). In the case of this particular model it is possible to show analytically that a topological change in the configuration space exists and that it can be related to the thermodynamic phase transition [104].

Let us consider again, as was already done in Section 6.2.1, the family M_v of submanifolds of the configuration space defined in Eq. (123); now the potential energy per degree of freedom is that of the mean-field XY model, i.e.,

$$\mathscr{V}(\varphi) = \frac{V(\varphi)}{N} = \frac{J}{2N^2} \sum_{i,j=1}^{N} [1 - \cos(\varphi_i - \varphi_j)] - \frac{h}{N} \sum_{i=1}^{N} \cos \varphi_i \,, \tag{134}$$

where $\varphi_i \in [0, 2\pi]$. Such a function can be considered a Morse function on M, so that, according to Morse theory (see Appendix B), all these manifolds have the same topology until a critical level $\mathscr{V}^{-1}(v_c)$ is crossed, where the topology of M_v changes.

A change in the topology of M_v can only occur when v passes through a critical value of $\mathscr{V}$. Thus, in order to detect topological changes in M_v we have to find the critical values of $\mathscr{V}$, which means solving the equations

$$\frac{\partial \mathscr{V}(\varphi)}{\partial \varphi_i} = 0, \qquad i = 1, \ldots, N \,. \tag{135}$$

For a general potential energy function $\mathscr{V}$, the solution of the Eqs. (135) would be a formidable task [111], but in the case of the mean-field XY model, the mean-field character of the interaction greatly simplifies the analysis, allowing an analytical treatment of the Eqs. (135); moreover, a projection of the configuration space onto a two-dimensional plane is possible.

We recall that in the limit $h \to 0$, the system has a continuous phase transition, with classical critical exponents, at $T_c = 1/2$, or $\varepsilon_c = 3/4$, where $\varepsilon = E/N$ is the energy per particle. We aim at showing that this phase transition has its foundation in a basic topological change that occurs in the configuration space M of the system. Let us remark that since $\mathscr{V}(\varphi)$ is bounded, $-h \le \mathscr{V}(\varphi) \le 1/2 + h^2/2$, the manifold is empty as long as $v < -h$, and when v increases beyond $1/2 + h^2/2$ no changes in its topology can occur so that the manifold M_v remains the same for any $v > 1/2 + h^2/2$, and is then an N-torus. To detect topological changes we have to solve Eqs. (135). To this end it is useful to define the magnetization vector, i.e., the collective spin vector $\boldsymbol{m} = (1/N)\sum_{i=1}^{N} \boldsymbol{s}_i$, which as a function of the angles is given by

$$\boldsymbol{m} = (m_x, m_y) = \left(\frac{1}{N} \sum_{i=1}^{N} \cos\varphi_i, \frac{1}{N} \sum_{i=1}^{N} \sin\varphi_i \right). \tag{136}$$

Due to the mean-field character of the model, the potential energy (116) can be expressed as a function of $\boldsymbol{m}$ alone (remember that $J = 1$), so that the potential energy per particle reads

$$\mathscr{V}(\varphi) = \mathscr{V}(m_x, m_y) = \tfrac{1}{2}(1 - m_x^2 - m_y^2) - h\, m_x\,. \tag{137}$$

This allows us to write the Eqs. (135) in the form ($i = 1, \ldots, N$)

$$(m_x + h)\sin\varphi_i - m_y \cos\varphi_i = 0\,. \tag{138}$$

Now we can solve these equations and find all the critical values of $\mathscr{V}$. The solutions of Eqs. (138) can be grouped in three classes:

(i) The minimal energy configuration $\varphi_i = 0\ \forall i$, with a critical value $v = v_0 = -h$, which tends to 0 as $h \to 0$. In this case, $m_x^2 + m_y^2 = 1$.

(ii) Configurations such that $m_y = 0, \sin\varphi_i = 0\ \forall i$. These are the configurations in which φ_i equals either 0 or π; i.e., we have again $\varphi_i = 0\ \forall i$, but also the N configurations with $\varphi_k = \pi$ and $\varphi_i = 0\ \forall i \neq k$, as well as the $N(N-1)$ configurations with two angles equal to π and all the others equal to 0, and so on, up to the configuration with $\varphi_i = \pi\ \forall i$. The critical values corresponding to these critical points depend only on the number of π's, n_π, so that $v(n_\pi) = \frac{1}{2}[1 - (1/N^2)(N - 2n_\pi)^2] - (h/N)(N - 2n_\pi)$. We see that the largest critical value is, for N even, $v(n_\pi = N/2) = 1/2$ and that the number of critical points corresponding to it is $\mathscr{O}(2^N)$.

(iii) Configurations such that $m_x = -h$ and $m_y = 0$, which correspond to the critical value $v_c = 1/2 + h^2/2$, which tends to $1/2$ as $h \to 0$. The number of these configurations grows with N not slower than $N!$ [104].

Configurations (i) are the absolute minima of $\mathscr{V}$, (iii) are the absolute maxima, and (ii) are all the other stationary configurations of $\mathscr{V}$.

Since for $v < v_0$ the manifold is empty, the topological change that occurs at v_0 is the one corresponding to the "birth" of the manifold from the empty set; subsequently there are many topological changes at values $v(n_\pi) \in (v_0, 1/2]$ till at v_c there is a final topological change which

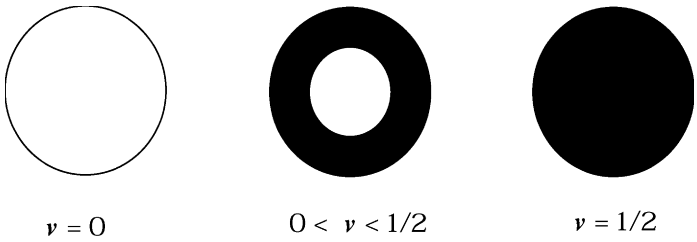

Fig. 27. The sequence of topological changes undergone by the manifolds D_v with increasing v in the limit $h \to 0$.

corresponds to the "completion" of the manifold. We remark that the number of critical values in the interval $[v_0,1/2]$ grows with N and that eventually the set of these critical values becomes dense in the limit $N \to \infty$. However, the critical value v_c remains isolated from other critical values also in that limit. We observe that it is necessary to consider a non-zero external field h in order that $\mathscr{V}$ is a Morse function, because if $h = 0$ all the critical points of classes (i) and (ii) are degenerate, in which case topological changes do not necessarily occur.[24] This degeneracy is due to the O(2)-invariance of the potential energy in the absence of an external field. To be sure, for $h \neq 0$, $\mathscr{V}$ may not be a Morse function on the whole of M either, but only on M_v with $v < v_c$, because the critical points of class (iii) may also be degenerate, so that v_c does not necessarily correspond to a topological change. However, this difficulty could be dealt with by using that the potential energy can be written in terms of the collective variables m_x and m_y – as in Eq. (137). This implies that we consider the system of N spins projected onto the two-dimensional configuration space of the collective spin variables. According to definition (136) of $\boldsymbol{m}$, the accessible configuration space is now not the whole plane, but only the disk

$$D = \{(m_x, m_y)\colon m_x^2 + m_y^2 \leq 1\}\,. \tag{139}$$

Thus, we want to study the topology of the submanifolds

$$D_v = \{(m_x, m_y) \in D\colon \mathscr{V}(m_x, m_y) \leq v\}\,. \tag{140}$$

The sequence of topological transformations undergone by D_v can now be very simply determined in the limit $h \to 0$ (see Fig. 27), as follows. As long as $v < 0$, D_v is the empty set. The first topological change occurs at $v = v_0 = 0$ where the manifold appears as the circle $m_x^2 + m_y^2 = 1$, i.e., the boundary ∂D of D. Then as v grows D_v is given by the conditions

$$1 - 2v \leq m_x^2 + m_y^2 \leq 1\,, \tag{141}$$

i.e., it is the ring with a hole centered at $(0,0)$ (punctuated disk) comprised between two circles of radii 1 and $\sqrt{2v}$, respectively. As v continues to grow the hole shrinks and is eventually completely filled as $v = v_c = 1/2$, where the second topological change occurs. In this coarse-grained

[24] It would also be possible to avoid this problem by considering an improved version of Morse theory, referred to as equivariant Morse theory [112].

two-dimensional description in D, all the topological changes that occur in M between $v = 0$ and $1/2$ disappear, and only the two topological changes corresponding to the extrema of $\mathscr{V}$, occurring at $v = v_0$ and v_c, survive. This means that the topological change at v_c should be present also in the full N-dimensional configuration space, so that the degeneracies mentioned above for the critical points of class (iii) should not prevent a topological change.

Now, we want to argue that the topological change occurring at v_c is related to the thermodynamic phase transition of the mean-field XY model. Since the Hamiltonian is of the standard form (1), the temperature T, the energy per particle ε and the average potential energy per particle $u = \langle \mathscr{V} \rangle$ obey, in the thermodynamic limit, the following equation:

$$\varepsilon = \frac{T}{2} + u(T), \tag{142}$$

where we have set Boltzmann's constant equal to 1. Substituting the values of the critical energy per particle $\varepsilon_c = 3/4$ and of the critical temperature $T_c = 1/2$ we get $u_c = u(T_c) = 1/2$, so that the critical value of the potential energy per particle v_c where the last topological change occurs equals the statistical-mechanical average value of the potential energy at the phase transition,

$$v_c = u_c\,. \tag{143}$$

Thus although a topological change in M occurs at any N, and v_c is *independent* of N, there is a connection of such a topological change and a thermodynamic phase transition *only* in the limit $N \to \infty$, $h \to 0^+$, when indeed a thermodynamic phase transition can be defined.

A similar kind of difference, as here between topological changes in mathematics (for all N) and phase transitions in physics (for $N \to \infty$ only), also occurs in other contexts in statistical mechanics, e.g. in non-equilibrium stationary states [113].

Since *not all* topological changes correspond to phase transitions, those that do correspond, remain to be determined to make the conjecture of Ref. [36] more precise. In this context, we consider one example where there are topological changes very similar to the ones of our model but no phase transitions, i.e., the one-dimensional XY model with nearest-neighbour interactions, whose Hamiltonian is of class (1) with interaction potential

$$V(\varphi) = \frac{1}{4}\sum_{i=1}^{N} [1 - \cos(\varphi_{i+1} - \varphi_i)] - h\sum_{i=1}^{N} \cos\varphi_i\,. \tag{144}$$

In this case the configuration space M is still an N-torus, and using again the potential energy per degree of freedom $\mathscr{V} = V/N$ as a Morse function, we can see that also here there are many topological changes in the submanifolds M_v as v is varied in the interval $[0, 1/2]$ (after taking $h \to 0^+$). However, there are critical points of the type $\varphi_j = \varphi_k = \varphi_l = \cdots = \pi$, $\varphi_i = 0\ \forall i \neq j, k, l, \ldots$; at variance with the mean-field XY model, it is now no longer the number of π's that determines the value of $\mathscr{V}$ at the critical point, but rather the number of domain walls, n_d, i.e., the number of boundaries between "islands" of π's and "islands" of 0's: $v(n_d) = n_d/2N$. Since $n_d \in [0, N]$, the critical values lie in the same interval as in the case of the mean-field XY model; but now the maximum critical value $v = 1/2$, instead of corresponding to a huge number of critical points, which rapidly grows with N, corresponds to *only two* configurations with N domain walls, which are $\varphi_{2k} = 0$, $\varphi_{2k+1} = \pi$, with $k = 1, \ldots, N/2$, and the reversed one. There are also

"spin-wave-like" critical points, i.e., such that $e^{i\theta_k} = \text{const}\, e^{2\pi ikn/N}$ with $n = 1, \ldots, N$ [114]; their critical energies are comprised in the interval above but again there is not a critical value associated to a huge number of critical points.

Thus, this example suggests the conjecture that a topological change in the configuration space submanifolds M_v occurring at a critical value v_c, is associated with a phase transition in the thermodynamic limit only if the number of critical points corresponding to the critical value v_c is sufficiently rapidly growing with N. On the basis of the behaviour of the mean-field XY model we expect then that such a growth should be at least exponential. Furthermore, a relevant feature appears to be that v_c remains an isolated critical value also in the limit $N \to \infty$: in the mean-field XY model this holds only if the thermodynamic limit is taken *before* the $h \to 0^+$ limit: this appears as a topological counterpart of the non-commutativity of the limits $h \to 0^+$ and $N \to \infty$ in order to get a phase transition in statistical mechanics.

The sequence of topological changes occurring with growing $\mathscr{V}$ makes the configuration space larger and larger, till at v_c the whole configuration space becomes fully accessible to the system through the last topological change. From a physical point of view, this corresponds to the appearance of more and more disordered configurations as T grows, which ultimately lead to the phase transition at T_c. We remark that the connection between the topology of the configuration space and the physics of continuous phase transitions made here via the potential energy, in particular Eq. (143), only makes sense in the thermodynamic limit, where the potential energy per particle $u(T)$ is well defined since its fluctuations vanish then at least as $1/\sqrt{N}$. This holds for our mean-field model, since for such a model, all fluctuations are absent. In the case of a real continuous (critical) phase transition the non-trivial role of fluctuations may complicate the present picture.

6.3. The topological hypothesis

The statistical behaviour of physical systems described by natural Hamiltonians is obtained, in the canonical ensemble, from the partition function in phase space

$$Z_N(\beta) = \int \mathrm{d}p\, \mathrm{d}q\, \mathrm{e}^{-\beta H(p,q)} = \left(\frac{\pi}{\beta}\right)^{N/2} \int \mathrm{d}q\, \mathrm{e}^{-\beta V(q)}$$

$$= \left(\frac{\pi}{\beta}\right)^{N/2} \int_0^\infty \mathrm{d}u\, \mathrm{e}^{-\beta u} \int_{\Sigma_u} \frac{\mathrm{d}\sigma}{\|\nabla V\|}, \tag{145}$$

where $p = (p_1, \ldots, p_N)$, $\mathrm{d}p = \prod_{i=1}^N \mathrm{d}p_i, q = (q_1, \ldots, q_N), \mathrm{d}q = \prod_{i=1}^N \mathrm{d}q_i$; $\Sigma_u = \{(q_1, \ldots, q_N) \in \boldsymbol{R}^N | V(q_1, \ldots, q_N) = u\}$ are equipotential hypersurfaces of configuration space and $\mathrm{d}\sigma$ is the measure on Σ_u induced from $\boldsymbol{R}^N$. The last term of the equation above shows that for Hamiltonians (1) the relevant statistical information is contained in the canonical configurational partition function $Z_N^C = \int \mathrm{d}q \exp[-\beta V(q)]$. Moreover, the last term of Eq. (145), written using a co-area formula [108], decomposes Z_N^C into an infinite summation of geometric integrals $\int_{\Sigma_u} \mathrm{d}\sigma/\|\nabla V\|$. This decomposition provides the point of attack for the formulation of a general hypothesis about a deep link between geometry, topology and thermodynamics and, obviously, phase transitions. In fact, once the potential energy $V(q)$ is given, the configuration space of the system is automatically

foliated into the family $\{\Sigma_u\}_{u\in R}$ of these equipotential hypersurfaces. Now, from standard statistical mechanical arguments concerning the equivalence of canonical and microcanonical ensembles we know that, at any given value of the inverse temperature β, the larger the number N of particles, the closer to some $\Sigma_{\bar{u}}$ are the microstates that significantly contribute to the statistical averages of thermodynamic observables. The hypersurface $\Sigma_{\bar{u}}$ is the one associated with $\bar{u} \equiv u_\beta = (Z_N^C)^{-1}\int\prod \mathrm{d}q_i\, V(q)\mathrm{e}^{-\beta V(q)}$, the average potential energy computed at a given β. Thus, at any β, if N is very large the effective support of the canonical measure shrinks very close to a single $\Sigma_v = \Sigma_{u_\beta}$. On the basis of this and what was found in [36,104,102], we formulate the following.

Topological hypothesis: *The basic origin of a phase transition lies in a topological change of the support of the measure describing a system. This change of topology induces a change of the measure itself at the transition point.*

This hypothesis stipulates that some change of the topology of the $\{\Sigma_u\}$, occurring at some $u_c = u_c(\beta_c)$, could be the origin of the singular behaviour of thermodynamic observables at a phase transition rather than measure singularities which in this view are induced from a deeper level where the topology changes take place.

In other words, the claim is that the canonical measure should "feel" a big and sudden change – if any – of the topology of the equipotential hypersurfaces of its underlying support, with as a consequence, the appearance of the typical signals of a phase transition, i.e. almost singular (at finite N) energy or temperature dependences of the averages of appropriate observables. The larger N, the narrower the effective support is of the measure – as discussed above – and hence the sharper the mentioned signals can be, until true singularities appear in the $N \to \infty$ limit.

We emphasize though that not all topological transitions lead to physical phase transitions. At present the precise connection between topological transitions and phase transitions still has to be clarified in many respects. Certain is that not every topological transition corresponds to a phase transition, as has been discussed in Sections 6.2.2 and 6.2.3. Rather it seems that, on the basis of present evidence, a phase transition corresponds to a super-combination of many simultaneous topological transitions taking place, where many might mean at least exponentially growing with the number of particles. It seems therefore more like a supertopologically constructed transition. This is illustrated analytically by the above discussed XY mean-field model, where an exponential crowding of topological transitions occurs on one side of the phase transition. Though such an analysis has not been possible for a numerically treated lattice φ^4 model, on the other hand, like the Euler characteristic χ clearly shows, a phase transition corresponds to an abrupt transition between different rates of change in the topology above and below the phase transition point; no such information is available for the analytic XY mean-field model where the calculation of χ has not yet been done.

6.4. Open questions and future developments

1. The phase space trajectories of dynamical systems described by Hamiltonian functions of the form (1), i.e.,

$$\mathscr{H} = \frac{1}{2}\sum_{i=1}^{N}\pi_i^2 + V(\varphi),$$

where the φ's and the π's are, respectively, the coordinates and the conjugate momenta, are bound to the constant energy hypersurfaces Σ_E of the $2N$-dimensional phase space spanned by the φ's and the π's. Therefore, it would be natural to investigate the relationship between the topology variations of these hypersurfaces in phase space and phase transitions. The Hamiltonian $\mathscr{H}$ would be the Morse function in this case. Moreover, having already considered the role played by the Σ_u of configuration space, we can wonder what is the relationship between topology changes in configuration space and phase space, respectively. Such a relationship is somehow subtle. At first sight all the critical points of configuration space are embodied also in critical points of phase space, in fact all the critical points of $\mathscr{H}$ are such that, $\forall i$, $\pi_i = 0$ and $\nabla_i V(\varphi) = 0$. However, the critical points of $\mathscr{H}$ physically correspond to vanishing kinetic energy, so that, if those topology changes of the Σ_u that are associated to a phase transition were to correspond to topology changes of the Σ_E – because the critical points of $\mathscr{H}$ incorporate those of $V(\varphi)$ – then the critical potential energy density and the critical total energy density at a phase transition should coincide, which is not the case. An argument to clarify this point can be given [50] by observing that the dynamics does not equally sample the whole Σ_E: the larger N, the smaller the relative fluctuations of the potential energy $\langle\delta^2 V\rangle^{1/2}/\langle V\rangle$ and kinetic energy $\langle\delta^2 K\rangle^{1/2}/\langle K\rangle$, respectively, are. Thus, by putting $v \equiv \langle V\rangle$ and $t \equiv \langle K\rangle$, we can assume that, at large N and given energy E, the momenta mainly live close to the hypersphere $\boldsymbol{S}_t^{N-1} = \{(p_1, \ldots, p_N) \in \boldsymbol{R}^N \,|\, \sum_{i=1}^N \frac{1}{2} p_i^2 = t\}$ and the Lagrangian coordinates mainly live close to the equipotential hypersurface Σ_v^{N-1} with $v + t = E$. Therefore, though the microcanonical measure mathematically extends over a whole energy surface, as far as physics is concerned, at very large N a non-negligible contribution to the microcanonical measure is, in practice, given only by a small subset of an energy surface. This subset can be approximately modelled by the product manifold $\Sigma_v^{N-1} \times \boldsymbol{S}_t^{N-1}$. Since the kinetic energy submanifold $\boldsymbol{S}_t^{N-1}$ is a hypersphere at any t, a change in the topology of Σ_v^{N-1} directly entails a change of the topology of $\Sigma_v^{N-1} \times \boldsymbol{S}_t^{N-1}$, that is of the effective model-manifold for the subset of Σ_E where the dynamics mainly "lives" at a given energy E.

This confirms that, as long as we are interested only in classical Hamiltonian systems of the standard form (1), we can restrict our geometrical and topological investigation to the submanifolds of the N-dimensional configuration space M.

2. A recent and important advance has been achieved at a more mathematical level. In Refs. [103,115,116] a theorem has been proved that establishes the *necessity* of topological changes of the equipotential hypersurfaces Σ_u for the appearance of first- or second-order thermodynamic phase transitions. The theorem applies to a wide class of finite-range potentials, bound below, describing systems confined in finite regions of space with continuously varying coordinates. The proof proceeds by showing that, under the crucial assumption of diffeomorphicity of the Σ_u in an arbitrary interval of values for u, the Helmholtz free energy is uniformly convergent in N to its thermodynamic limit, at least within the class of twice differentiable functions, in a corresponding interval of temperature.

This theorem confirms the general validity of the TH and ensures that for a wide class of physical potentials the mathematical framework of differential topology is adequate to describe, at least, first- and second-order phase transitions. There is no proof of *sufficiency*. On the basis of the discussions in Sections 6.2.2, 6.2.3 and 6.3 we already know that a simple loss of diffeomorphicity of the Σ_u is not sufficient to lead to a phase transition.

A thorough investigation of those classes of topology changes that are responsible for the appearance of phase transitions is at present the main challenge of this new point of view about phase transitions and certainly represents a topic that will remunerate the efforts addressed to it.

3. Let us finally highlight some interesting related topics.

(i) We might speculate about the possibility of relating universal quantities of the theory of critical phenomena, like critical exponents, to some topological counterpart; in fact a notion of universality arises quite naturally in a topological framework.
(ii) Topology provides a common ground for the roots of both dynamics and thermodynamics: insofar as the dynamics of the system, i.e. the motion of the trajectory in phase space, takes place in what was called before the support of the statistical measure, it is clear that the nature of the trajectory will crucially depend on the topology of the manifold to which it belongs. This therefore strenghtens the interest of a dynamical treatment of phase transitions giving new emphasis to the microcanonical ensemble and thereby joining other recent developments in the field [117]. Since the dynamical approach does not depend on whether a system is in statistical equilibrium, non-equilibrium, or in a metastable state (like a glass and, more generally, amorphous materials) a dynamical approach to phase transitions might also be important for systems whose thermodynamical state is not well defined.
(iii) The fact that the topological changes appear at any N opens a new possibility to study transitional phenomena in *finite* systems, like nuclear and atomic clusters, polymers and proteins, or other biological systems, as well as for nano and mesoscopic structures.

Acknowledgements

We want to dedicate this paper to the memory of our friend and brilliant collaborator Lando Caiani, prematurely deceased, who gave seminal contributions to many aspects of the approach described here, and especially to the early development of the connection between topology and phase transitions.

During their development, the main ideas described in the present Report have been discussed with many colleagues, among which we would like to thank, for their particular interest and for many useful suggestions and criticism, and sometimes for an active collaboration, S. Caracciolo, M. Cerruti-Sola, C. Clementi, Y. Elskens, M.-C. Firpo, R. Franzosi, R. Gatto, R. Livi, G. Mussardo, H.A. Posch, V. Penna, G. Pettini, M. Rasetti, S. Ruffo, L. Spinelli, G. Vezzosi, A.M. Vinogradov, and the DOCS research group in Firenze (http://docs.de.unifi.it/ ~ docs/).

Part of this work has been done at the Erwin Schrödinger Institut für Mathematische Physik (ESI) in Vienna, whose kind hospitality is gratefully acknowledged. LC thanks the Rockefeller University in New York for its kind hospitality and for partial financial support, the 1997–1999 SINTESI program of the Italian Ministry of University and of Scientific and Technologic Research (MURST) for partial financial support, and Prof. H.A. Posch for hospitality at the Institut für Experimental Physik of the University of Vienna. EGDC is indebted to the Engineering Research Program of the Office of Basic Engineering Section of the US Department of Energy for support under grant DE-FG02-88-ER13847, as well as to the Politecnico di Torino and to the INFM, UdR Torino Politecnico, for hospitality and partial financial support.

This work has been partially supported by the INFM, section G, under the PAIS "Equilibrium and Non-Equilibrium Dynamics in Condensed Matter Systems".

Appendix A. Summary of Riemannian geometry

In the following, we briefly recall some essential concepts and notations of Riemannian differential geometry which are used in the main text. The present section is only meant to facilitate the reader to follow the main text of the Report, so that our discussion will not be a rigorous treatment of the subject. For a more elaborate discussion, we refer the reader to a textbook of general relativity (e.g., the classic Landau and Lifshitz's book [118] or the more recent and complete, but still very clear and readable, textbook by Wald [119]). A more mathematically oriented, but still readable by physicists, introduction to the subject is given by do Carmo [59]; a comprehensive and rigorous treatment, which, however, goes far beyond what is needed to follow the exposition in the main text, can be found in Kobayashi and Nomizu [120].

The Einstein summation convention over repeated indices is always understood unless explicitly stated to the contrary. Moreover, we follow throughout the paper the usual convention to suppress the dependence of the components of vector and tensor quantities on the (proper) time and, in general, only indicate it explicitly when this dependence is absolutely relevant.

A.1. Riemannian manifolds

A set M is called a *differentiable manifold* if it can be covered with a collection, either finite or denumerable, of *charts*, such that each point of M is represented at least on one chart, and the different charts are differentiably connected to each other. A chart is a set of coordinates on the manifold, i.e., it is a set of n real numbers $(x^1, \ldots, x^n)$ which denote the "position" of a point on the manifold. The number n of coordinates of a chart is the same for each connected part of the manifold (and for the whole manifold if the latter is connected, i.e., it cannot be split in two disjoint parts which are still manifolds); such a number is called the *dimension* of the manifold M. The union of the charts on M is called an *atlas* of M.

A.1.1. Vectors and tensors

A vector (more precisely, a *tangent* vector), can be defined using curves on the manifold M. Given a curve γ in M, represented in local coordinates by the parametric equations $x = \varphi(t)$, we define a tangent vector at $P \in M$ as the velocity vector of the curve in P, i.e.,

$$v = \dot{\gamma} = \lim_{t \to 0} \frac{\varphi(t) - \varphi(0)}{t}, \quad \varphi(0) = P, \tag{A.1}$$

so that the n components of the tangent vector v are given by

$$v^i = \frac{\mathrm{d}\varphi^i}{\mathrm{d}t}. \tag{A.2}$$

The set of all the tangent vectors of M in P is a linear space, referred to as the *tangent space* of M in P, and denoted by $T_P M$. Each tangent space is isomorphic to an n-dimensional Euclidean space.

Given a chart $(x^1,\dots,x^n)$ in a neighborhood of P, a basis $(X_1,\dots,X_n)$ of T_PM can be defined, so that a generic vector v is expressed as a sum of the X_i's weighted by its components,

$$v = v^i X_i\,. \tag{A.3}$$

The basis $\{X_i\}$ is called a *coordinate basis* of T_PM, and its components X_i are often denoted[25] by $\partial/\partial x^i$. The basis depends on the chart: choosing another chart, $(x^{1'},\dots,x^{n'})$, we get another basis $\{X_i'\}$. The components of v in the two different bases are connected by the following rule:

$$v^{i'} = v^j \frac{\partial x^{i'}}{\partial x^j}\,, \tag{A.4}$$

referred to as the *vector transformation rule*. Indeed, one can define a vector as a quantity whose components transform according to Eq. (A.4). The union of all the tangent spaces of the manifold M,

$$TM = \bigcup_{P\in M} T_PM \tag{A.5}$$

is a $2n$-dimensional manifold and is referred to as the *tangent bundle* of M.

A *vector field* V on M is an assignment of a vector v_P at each point $P\in M$. If f is a smooth function,

$$V(f)|_P = v_P(f) \tag{A.6}$$

is a real number for each $P\in M$, i.e., $v(f)$ is a function on M. If such a function is smooth, V is called a *smooth vector field* on M. The curves $\varphi(t)$ which satify the differential equations

$$\dot\varphi = V(\varphi(t)) \tag{A.7}$$

are called the *trajectories* of the field V, and the mapping $\varphi_t : M\mapsto M$ which maps any point P of M along the trajectory of V emanating from P is called the *flow* of V. Given two vector fields V, W, one can define the *commutator* as the vector field $[V,W]$ such that

$$[V,W](f) = V(W(f)) - W(V(f))\,, \tag{A.8}$$

i.e., in terms of the local components,

$$[V,W]^j = V^i\frac{\partial W^j}{\partial x^i} - W^i\frac{\partial V^j}{\partial x^i}\,. \tag{A.9}$$

We note that, if $\{X_i\}$ is a coordinate basis,

$$[X_i, X_j] = 0 \quad \forall i,j \tag{A.10}$$

and that, conversely, given n non-vanishing and commuting vector fields which are linearly independent, there always exists a chart for which these vector fields are a coordinate basis.

Tangent vectors are not the only vector-like quantities that can be defined on a manifold M: there are also *cotangent* vectors, which can be defined as follows. Let us recall that the *dual space*

[25] The origin of this notation is in the fact that vectors can be defined as directional derivatives on M (see e.g. Ref. [119]).

V^* of a vector space V is the space of *linear* maps from V to the real numbers. Given a basis of V, $\{u_i\}$, a basis of V^*, $\{u^{i*}\}$, called the *dual basis*, is defined by

$$u^{i*}(u_j) = \delta^i_j\,. \tag{A.11}$$

The dual space of TM, T^*M, is called the *cotangent bundle* of M. Its elements are called *cotangent vectors*, or sometimes *covariant vectors* (while the tangent vectors are sometimes denoted as *contravariant* vectors). The dual basis elements are usually denoted as $\mathrm{d}x^1, \ldots, \mathrm{d}x^n$, i.e., $\mathrm{d}x^i$ is such that $\mathrm{d}x^i(\partial/\partial x^j) = \delta^i_j$. The components ω_i of cotangent vectors transform according to the rule

$$\omega'_i = \omega_j \frac{\partial x^j}{\partial x'^i} \tag{A.12}$$

to be compared with Eq. (A.4). The common rule is to use subscripts to denote the components of dual vectors and superscripts for those of vectors.

A (k,l)-*tensor* T over a vector space V is a multilinear map

$$T : \underbrace{V^* \times \cdots \times V^*}_{k \text{ times}} \times \underbrace{V \times \cdots \times V}_{l \text{ times}} \mapsto \boldsymbol{R}\,, \tag{A.13}$$

i.e., acting on k dual vectors and l vectors, T yields a number, and it does so in such a manner that if we fix all but one of the vectors or dual vectors, it is a linear map in the remaining variable. A (0,0) tensor is a scalar, a $(0,1)$ tensor is a vector, and a $(1,0)$ tensor is a dual vector. The space $\mathscr{T}(k,l)$ of the tensors of type (k,l) is a linear space; a (k,l)-tensor is defined once its action on k vectors of the dual basis and on l vectors of the basis is known, and since there are $n^k n^l$ independent ways of choosing these basis vectors, $\mathscr{T}(k,l)$ is a n^{k+l}-dimensional linear space. Two natural operations can be defined on tensors. The first one is called *contraction* with respect to the ith (dual vector) and the jth (vector) arguments and is a map

$$C : T \in \mathscr{T}(k,l) \mapsto CT \in \mathscr{T}(k-1, l-1) \tag{A.14}$$

defined by

$$CT = \sum_{\sigma=1}^{n} T(\ldots, \underbrace{v^{\sigma *}}_{i}, \ldots; \ldots, \underbrace{v^{\sigma}}_{j}, \ldots)\,. \tag{A.15}$$

The contracted tensor CT is independent of the choice of the basis, so that the contraction is a well defined, invariant, operation. The second operation is the *tensor product*, which maps an element $\mathscr{T}(k,l) \times \mathscr{T}(k',l')$ into an element of $\mathscr{T}(k+k', l+l')$, i.e., two tensors T and T' into a new tensor, denoted by $T \otimes T'$, defined as follows: given $k+k'$ dual vectors $v^{1^*}, \ldots, v^{k+k'^*}$ and $l+l'$ vectors $w_1, \ldots, w_{l+l'}$, then

$$\begin{aligned} T \otimes T'(v^{1^*}, \ldots, v^{k+k'^*}; w_1, \ldots, w_{l+l'}) &= T(v^{1^*}, \ldots, v^{k^*}; w_1, \ldots, w_l) \\ &\quad T'(v^{k+1^*}, \ldots, v^{k+k'^*}; w_{l+1}, \ldots, w_{l+l'})\,. \end{aligned} \tag{A.16}$$

The tensor product allows one to construct a basis for $\mathscr{T}(k,l)$ starting from a basis $\{v_\mu\}$ of V and its dual basis $\{v^{\nu *}\}$: such a basis is given by the n^{k+l} tensors $\{v_{\mu_1} \otimes \cdots \otimes v_{\mu_k} \otimes v^{\nu_1 *} \otimes \cdots \otimes v^{\nu_l *}\}$. Thus, every tensor $T \in \mathscr{T}(k,l)$ allows a decomposition

$$T = \sum_{\mu_1, \ldots, \nu_l = 1}^{n} T^{\mu_1 \cdots \mu_k}{}_{\nu_1 \cdots \nu_l} v_{\mu_1} \otimes \cdots \otimes v^{\nu_l *}\,, \tag{A.17}$$

the numbers $T^{\mu_1 \cdots \mu_k}{}_{\nu_1 \cdots \nu_l}$ are called the *components* of T in the basis $\{v_\mu\}$. The components of the contracted tensor CT are

$$(CT)^{\mu_1 \cdots \mu_{k-1}}{}_{\nu_1 \cdots \nu_{l-1}} = T^{\mu_1 \cdots \sigma \cdots \mu_k}{}_{\nu_1 \cdots \sigma \cdots \nu_l} \tag{A.18}$$

(remember the summation convention), and the components of the tensor product $T \otimes T'$ are

$$(T \otimes T')^{\mu_1 \cdots \mu_{k+k'}}{}_{\nu_1 \cdots \nu_{l+l'}} = T^{\mu_1 \cdots \mu_k}{}_{\nu_1 \cdots \nu_l} T'^{\mu_{k+1} \cdots \mu_{k+k'}}{}_{\nu_{l+1} \cdots \nu_{l+l'}} . \tag{A.19}$$

All these results are valid for a generic vector space, so that they hold, in particular, for the vector spaces of the tangent bundle TM of M, over which tensors (and, analogously to vector fields, *tensor fields*) can be defined exactly as above.

A.1.2. Riemannian metrics

The infinitesimal square distance on M, i.e., the length element $\mathrm{d}s^2$ (also referred to as the *metric*) can be defined at each point $P \in M$ by means of a $(0,2)$-tensor g, provided it is *symmetric*, i.e., $g(v,w) = g(w,v)$, and *non-degenerate*, i.e., $g(v,w) = 0 \ \forall v \in T_P M$ if and only if $w = 0$.[26] In fact, a g with these properties induces on the tangent bundle TM a nondegenerate quadratic form (called the *scalar product*),

$$g(v,w) = \langle v, w \rangle : TM \times TM \mapsto \boldsymbol{R} . \tag{A.20}$$

Then it is possible to measure lengths on the manifold. A manifold M, equipped with a scalar product, is called a (pseudo)*Riemannian manifold*, and the scalar product is referred to as a (pseudo)*Riemannian structure* on M. If the quadratic form (A.20) is positive definite, then one speaks of a (proper) *Riemannian metric*. In the latter case the squared length element is always positive. For instance, one can define the length of a curve as

$$\ell(\gamma) = \int_\gamma \sqrt{\langle \dot{\gamma}, \dot{\gamma} \rangle}\, \mathrm{d}t . \tag{A.21}$$

The curves γ which are extremals of the length functional are called the *geodesics* of M.

In a coordinate basis, we can expand the metric g as

$$g = g_{ij}\, \mathrm{d}x^i \otimes \mathrm{d}x^j , \tag{A.22}$$

so that one defines the invariant (squared) length element on the manifold, in local coordinates, as

$$\mathrm{d}s^2 = g_{ij}\, \mathrm{d}x^i \mathrm{d}x^j . \tag{A.23}$$

The scalar product of two vectors v and w is given, in terms of g, by

$$\langle v, w \rangle = g_{ij} v^i w^j = v_j w^j = v^i w_i . \tag{A.24}$$

In the above equation we have made use of the fact that g estabilishes a one-to-one correspondence between vectors and dual vectors, i.e., in components,

$$g_{ij} v^j = v_i . \tag{A.25}$$

[26] Or, equivalently, $g(v,w) = 0 \ \forall w \in T_P M$ if and only if $v = 0$; the two statements are equivalent because g is a symmetric tensor.

For this reason, the components of the inverse metric g^{-1} are simply denoted by g^{ij}, instead of $(g^{-1})^{ij}$, and allow to pass from dual vector (covariant) components to vector (contravariant) components:

$$g^{ij}v_j = v^i . \tag{A.26}$$

This operation of raising and lowering the indices can be applied not only to vector, but also to tensor components. This allows us to pass from (k,l) tensor components to the corrseponding $(k+1,l-1)$ tensor components and vice versa. What does not change in the operation is the sum $k+l$ which is called the *rank* (or the *order*) of the tensor.

A.2. Covariant differentiation

The introduction of a differential calculus on a manifold which is not Euclidean is complicated by the fact that ordinary derivatives do not map vectors into vectors, i.e., the ordinary derivatives of the components of a vector w, $\mathrm{d}w^i/\mathrm{d}t$, taken for instance at a point P along a given curve $\gamma(t)$, are *not* the components of a vector in $T_P M$, because they do not transform according to rule (A.4). The geometric origin of this fact is that the *parallel transport* of a vector from a point P to a point Q on a non-Euclidean manifold *does* depends on the path chosen to join P and Q. Since in order to define the derivative of a vector at P, we have to move that vector from P to a neighboring point along a curve and then parallel transport it back to the original point in order to measure the difference, we need a definition of parallel transport to define a derivative; conversely, given a (consistent) derivative, i.e., a derivative which maps vectors into vectors, one could define the parallel transport by imposing that a vector is parallel transported along a curve if its derivative along the curve is zero. The two ways are conceptually equivalent: we follow the first way, by introducing the notion of a *connection* and then using it to define the derivative operator. Such a derivative will be referred to as the *covariant derivative*.

A (linear) *connection* on the manifold M is a map ∇ such that, given two vector fields[27] A and B, it yields a third field $\nabla_A B$ with the following properties:

1. $\nabla_A B$ is bilinear in A and B, i.e., $\nabla_A(\alpha B + \beta C) = \alpha\nabla_A B + \beta\nabla_A C$ and $\nabla_{\alpha A+\beta B}C = \alpha\nabla_A C + \beta\nabla_B C$;
2. $\nabla_{f(A)}B = f(\nabla_A B)$;
3. (Leibnitz rule) $\nabla_A f(B) = (\partial_A f)B + f(\nabla_A B)$, where ∂_A is the ordinary directional derivative in the direction of A.

The *parallel transport* of a vector V along a curve γ, whose tangent vector field is $\dot{\gamma}$, is then defined as the (unique) vector field $W(t) = W(\gamma(t))$ along $\gamma(t)$ such that

1. $W(0) = V$;
2. $\nabla_{\dot{\gamma}} W = 0$ along γ.

The notion of covariant derivative now immediately follows: the *covariant derivative* $\mathrm{D}V/\mathrm{d}t$ of V along γ is given by the vector field

$$\frac{\mathrm{D}V}{\mathrm{d}t} = \nabla_{\dot{\gamma}} V . \tag{A.27}$$

[27] One could also consider tensor fields, but for the sake of simplicity we define connections using vectors.

On the basis of Eq. (A.27), with a certain abuse of language, one often refers to $\nabla_X Y$ as the covariant derivative of Y along X, where X and Y are generic vector fields. Among all the possible linear connections, and given a metric g, there is one and only one which (i) is *symmetric*, i.e.,

$$\nabla_X Y - \nabla_Y X = [X, Y] \quad \forall X, Y \tag{A.28}$$

and (ii) conserves the scalar product, i.e., the scalar product of two *parallel* vector fields P and P' is constant along γ,

$$\frac{\mathrm{d}}{\mathrm{d}t}\langle P, P'\rangle \equiv 0\,. \tag{A.29}$$

Such a linear connection is obviously the natural one on a Riemannian manifold, and is referred to as the *Levi–Civita* (or *Riemannian*) connection. Whenever we refer to a *covariant derivative* without any specification, we mean the covariant derivative induced by the Riemannian connection.

The components of the Riemannian connection ∇ with respect to a coordinate basis $\{X_i\}$ are the *Christoffel symbols*, given by

$$\Gamma^i_{jk} = \langle\, \mathrm{d}q^i, \nabla_{X_j} X_k \rangle \tag{A.30}$$

and are given, in terms of the derivatives of the components of the metric, by the following formula:

$$\Gamma^i_{jk} = \tfrac{1}{2} g^{im}(\partial_j g_{km} + \partial_k g_{mj} - \partial_m g_{jk})\,, \tag{A.31}$$

where $\partial_i = \partial/\partial x^i$. The expression in local coordinates of the covariant derivative (A.27) of a vector field V is then

$$\frac{\mathrm{D}V^i}{\mathrm{d}t} = \frac{\mathrm{d}V^i}{\mathrm{d}t} + \Gamma^i_{jk}\frac{\mathrm{d}x^j}{\mathrm{d}t}V^k\,. \tag{A.32}$$

A.2.1. Geodesics

The *geodesics*, which were already defined as the curves of extremal length on the manifold, can also be defined as *self-parallel curves*, i.e., curves such that the tangent vector $\dot\gamma$ is always parallel transported. Thus, geodesics are the curves $\gamma(t)$ which satisfy the equation (referred to as the *geodesic equation*)

$$\frac{\mathrm{D}\dot\gamma}{\mathrm{d}t} = 0 \tag{A.33}$$

whose expression in local coordinates follws from Eq. (A.32), and is

$$\frac{\mathrm{d}^2 x^i}{\mathrm{d}t^2} + \Gamma^i_{jk}\frac{\mathrm{d}x^j}{\mathrm{d}t}\frac{\mathrm{d}x^k}{\mathrm{d}t} = 0\,. \tag{A.34}$$

Since the norm of the tangent vector $\dot\gamma$ of a geodesic is constant, $|\mathrm{d}\gamma/\mathrm{d}t| = c$, the arc length of a geodesic is proportional to the parameter

$$s(t) = \int_{t_1}^{t_2}\left|\frac{\mathrm{d}\gamma}{\mathrm{d}t}\right|\mathrm{d}t = c(t_2 - t_1)\,. \tag{A.35}$$

When the parameter is actually the arc length, i.e., $c = 1$, we say that the geodesic is *normalized*. Whenever we consider a geodesic, we assume it is normalized, if not explicitly stated otherwise. This means that the Eqs. (A.34) are nothing but the Euler–Lagrange equations for the length functional along a curve $\gamma(s)$ parametrized by the arc-length

$$\ell(\gamma) = \int_\gamma \mathrm{d}s\,. \tag{A.36}$$

Given a geodesic $\gamma(s)$ on M, there exists a unique vector field G on TM such that its trajectories are $(\gamma(s), \dot{\gamma}(s))$. Such a vector field is called the *geodesic field* and its flow the *geodesic flow* on M.

A.3. Curvature

The curvature of a Riemannian manifold (M, g) is – intuitively – a way of measuring how much the manifold deviates from being Euclidean. The *curvature tensor*, also known as the *Riemann–Christoffel tensor*, is a tensor of order 4 defined as

$$R(X, Y) = \nabla_X \nabla_Y - \nabla_Y \nabla_X - \nabla_{[X,Y]}\,, \tag{A.37}$$

where ∇ is the Levi–Civita connection of M. Observe that if $M = \boldsymbol{R}^N$, then $R(X, Y) = 0$ for all the pairs of tangent vectors X, Y, because of the commutativity of the ordinary derivatives. In addition, R measures the non-commutativity of the covariant derivative: in fact, if we choose a coordinate system $\{x_1, \ldots, x_n\}$, we have, since $[\partial/\partial x_i, \partial/\partial x_j] = 0$,

$$R\left(\frac{\partial}{\partial x_i}, \frac{\partial}{\partial x_j}\right) = \nabla_{\partial/\partial x_i} \nabla_{\partial/\partial x_j} - \nabla_{\partial/\partial x_j} \nabla_{\partial/\partial x_i}\,. \tag{A.38}$$

In local coordinates, the components of the Riemann curvature tensor (considered here as a $(1, 3)$-tensor) are given by

$$R^i{}_{jkl} = \frac{\partial \Gamma^i_{jl}}{\partial x^k} - \frac{\partial \Gamma^i_{kl}}{\partial x^j} + \Gamma^r_{jl}\Gamma^i_{kr} - \Gamma^r_{kl}\Gamma^i_{jr}\,. \tag{A.39}$$

Thus, given a metric g, the curvature R is uniquely defined. A manifold (M, g) is called *flat* when the curvature tensor vanishes.

Given a *positive* function f^2, a *conformal transformation* is the transformation

$$(M, g) \mapsto (M, \tilde{g}); \quad \tilde{g} = f^2 g\,. \tag{A.40}$$

Two Riemannian manifolds are said *conformally related* if they are linked by a conformal transformation. In particular, a manifold is (M, g) *conformally flat* if it is possible to find a conformal transformation which sends g into a flat metric. Conformally flat manifolds exhibit some remarkable simplifications for the calculation of the curvature tensor components (see e.g. [65]; an application is given in Section 3.3).

Closely related to the curvature tensor is the sectional – or Riemannian – curvature, which we define now. Let us consider two vectors $u, v \in T_P M$, and let us put

$$|u \wedge v| = (|u|^2|v|^2 - \langle u, v\rangle)^{1/2}\,, \tag{A.41}$$

which is the area of the two-dimensional parallelogram determined by u and v. If $|u \wedge v| \neq 0$ the vectors u, v span a two-dimensional subspace $\pi \subset T_P M$. We define the *sectional curvature* at the point P relative to π, as the quantity:

$$K(P;u,v) = K(P,\pi) = \frac{\langle R(v,u)u, v\rangle}{|u \wedge v|^2} \tag{A.42}$$

which can be shown to be independent of the choice of the two vectors $u, v \in \pi$. In local coordinates, Eq. (A.42) becomes

$$K(P;u,v) = R_{ijkl}\frac{u^i v^j u^k v^l}{|u \wedge v|^2}\,. \tag{A.43}$$

The knowledge of K for the $N(N-1)$ planes π spanned by a maximal set of linearly independent vectors completely determines R at P.

If $\dim(M) = 2$ then K coincides with the Gaussian curvature of the surface, i.e., with the product of the reciprocals of two curvature radii.

A manifold is called *isotropic* if $K(P,\pi)$ does not depend on the choice of the plane π. The remarkable result – Schur's theorem [59] – is that in this case K is also constant, i.e. it does not depend on the point P either.

Some "averages" of the sectional curvatures are very important. The *Ricci curvature* K_R at P in the direction v is defined as the sum of the sectional curvatures at P relative to the planes determined by v and the $N-1$ directions orthogonal to v, i.e., if $\{e_1, \ldots, e_{N-1}, v = e_N\}$ is an orthonormal basis of $T_P M$ and π_i is the plane spanned by v and e_i,

$$K_R(P,v) = \sum_{i=1}^{N-1} K(P,\pi_i)\,. \tag{A.44}$$

The *scalar curvature* $\mathscr{R}$ at P is the sum of the N Ricci curvatures at P,

$$\mathscr{R}(P) = \sum_{i=1}^{N} K_R(P,e_i)\,. \tag{A.45}$$

In terms of the components of the curvature tensor, such curvatures can be defined as follows (in the following formulae, we drop the dependence on P, because it is understood that the components are local quantities). We first define the *Ricci tensor* as the two-tensor whose components, R_{ij}, are obtained by contracting the first and the third indices of the Riemann tensor

$$R_{ij} = R^k{}_{ikj}\,, \tag{A.46}$$

then

$$K_R(v) = R_{ij}v^i v^j\,. \tag{A.47}$$

The right-hand side of Eq. (A.47) is called "saturation" of R_{ij} with v. The scalar curvature can be obtained as the trace of the Ricci tensor,

$$\mathscr{R} = R^i{}_i\,. \tag{A.48}$$

In the case of a *constant curvature* – or isotropic – manifold, the components of the Riemann curvature tensor have the remarkably simple form

$$R_{ijkl} = K(g_{ik}g_{jl} - g_{il}g_{jk}), \tag{A.49}$$

where K is the constant sectional curvature, so that the components of the Ricci tensor are

$$R_{ij} = Kg_{ij} \tag{A.50}$$

and all the above defined curvatures are constants, and are related by

$$K = \frac{1}{N-1}K_R = \frac{1}{N(N-1)}\mathscr{R}. \tag{A.51}$$

A.4. The Jacobi equation

In this subsection we give a derivation of the Jacobi equation, already introduced in the main text as Eq. (27). We will proceed as follows: first, we will define the geodesic separation vector field J, then we will show that the field J is actually a Jacobi field, i.e., obeys the Jacobi equation.

Let us define a *geodesic congruence* as a family of geodesics $\{\gamma_\tau(s) = \gamma(s,\tau) \,|\, \tau \in \boldsymbol{R}\}$ issuing from a neighbourhood $\mathscr{I}$ of a point of a manifold, smoothly dependent on the parameter τ, and let us fix a reference geodesic $\gamma(s,\tau_0)$. Denote then by $\dot{\gamma}(s)$ the vector field tangent to $\gamma(s,\tau_0)$ in s, i.e., the velocity vector field whose components are

$$\dot{\gamma}^i = \frac{\mathrm{d}x^i}{\mathrm{d}s} \tag{A.52}$$

and by $J(s)$ the vector field tangent in τ_0 to the curves $\gamma_s(\tau)$ for a fixed s, i.e., the vector field of components

$$J^i = \frac{\mathrm{d}x^i}{\mathrm{d}\tau}. \tag{A.53}$$

The field J will be referred to as the *geodesic separation field*, and measures the distance between nearby geodesics, as is shown in Fig. 28.

Let us now show that J is a Jacobi field. First of all, we notice that the field J commutes with $\dot{\gamma}$, i.e., $[\dot{\gamma}, J] = 0$. In fact, from the definition of the commutator (Eq. (A.9)) and from the definitions of J, Eq. (A.53), and of $\dot{\gamma}$, Eq. (A.52), we have

$$[\dot{\gamma}, J]^i = \dot{\gamma}^j \frac{\partial J^i}{\partial x^j} - J^j \frac{\partial \dot{\gamma}^i}{\partial x^j} = \frac{\partial x^j}{\partial s}\frac{\partial J^i}{\partial x^j} - \frac{\partial x^j}{\partial \tau}\frac{\partial \dot{\gamma}^i}{\partial x^j} = \frac{\partial J^i}{\partial s} - \frac{\partial \dot{\gamma}^i}{\partial \tau} \tag{A.54}$$

and using again Eqs. (A.53) and (A.52), we find that

$$\frac{\partial J^i}{\partial s} = \frac{\partial}{\partial s}\frac{\partial x^i}{\partial \tau} = \frac{\partial}{\partial \tau}\frac{\partial x^i}{\partial s} = \frac{\partial \dot{\gamma}^i}{\mathrm{d}\tau}, \tag{A.55}$$

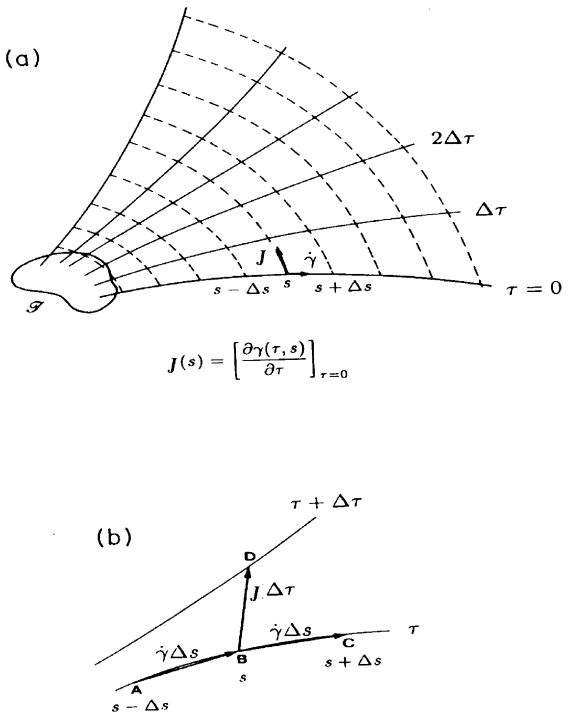

Fig. 28. Pictorial description of the definition of the geodesic separation vector J. From Ref. [22].

so that $[\dot\gamma, J] = 0$. Now, let us compute the second covariant derivative of the field J, $\nabla^2_{\dot\gamma} J$. First of all, let us recall that our covariant derivative comes from a Levi–Civita connection, which is symmetric (see Eq. (A.28)), so that

$$\nabla_{\dot\gamma} J - \nabla_J \dot\gamma = [\dot\gamma, J] \tag{A.56}$$

and having just shown that $[\dot\gamma, J] = 0$, we can write

$$\nabla_{\dot\gamma} J = \nabla_J \dot\gamma \tag{A.57}$$

Now, using this result, and the fact that $\nabla_{\dot\gamma}\dot\gamma = 0$ because $\bar\gamma$ is a geodesic, we can write

$$\nabla^2_{\dot\gamma} J = \nabla_{\dot\gamma}\nabla_{\dot\gamma} J = \nabla_{\dot\gamma}\nabla_J \dot\gamma = [\nabla_{\dot\gamma}, \nabla_J]\dot\gamma\,, \tag{A.58}$$

from which, using the definition of the curvature tensor (Eq. (A.37)) and, again, the vanishing of the commutator $[\dot\gamma, J]$, we get

$$\nabla^2_{\dot\gamma} J = R(\dot\gamma, J)\dot\gamma\,, \tag{A.59}$$

which is nothing but the Jacobi equation (27), written in compact notation.

It is worth noticing that the normal component $J_\perp$ of J, i.e., the component of J orthogonal to $\dot\gamma$ along the geodesic γ, is again a Jacobi field, since we can always write $J = J_\perp + \lambda\dot\gamma$: one immediately finds then that the velocity $\dot\gamma$ satisfies the Jacobi equation, so that $J_\perp$ must obey the

same equation. This allows us to restrict ourselves to the study of the normal Jacobi fields, as we have already done in the main text.

Appendix B. Summary of elementary Morse theory

The purpose of this Appendix is to recall the main ideas and concepts of Morse theory which are relevant for the main text of the paper. For a more elaborate discussion we refer the reader to Refs. [121–123].

Morse theory, also referred to as critical point theory, links the *topology* of a given manifold M with the properties of the *critical points* of smooth (i.e., with infinitely many derivatives) functions defined on it. Morse theory links *local* properties (what happens at a particular point of a manifold) with *global* properties (the topology, i.e., the shape, of the manifold as a whole). Two manifolds M and M' are topologically equivalent if they can be smoothly deformed one into the other: a tea cup is topologically equivalent to a doughnut, but it is *not* topologically equivalent to a ball. In fact a ball has no holes, while both a tea cup and a doughnut have one hole. To define precisely what a "smooth deformation" is, one has to resort to the notion of a *diffeomorphism*. A diffeomorphism is a smooth one-to-one map, whose inverse is smooth. Then M can be smoothly deformed into M' if there exists a diffeomorphism ψ which maps M into $M' = \psi(M)$. If such a diffeomorphism exists, we say that M and M' are then *diffeomorphic*. Thus the notion of "topological equivalence" we referred to has now a precise meaning.

For the sake of simplicity, we shall consider only compact, finite-dimensional manifolds: most of the results can be extended not only to noncompact manifolds, but also to infinite-dimensional manifolds modeled on Hilbert spaces (see [123]).

The key ingredient of Morse theory is to look at the manifold M as decomposed in the *level sets* of a function f. Let us recall that the a-level set of a function $f: M \mapsto \boldsymbol{R}$ is the set

$$f^{-1}(a) = \{x \in M : f(x) = a\}, \tag{B.1}$$

i.e., the set of all the points $x \in M$ such that $f(x) = a$. Now, M being compact, any function f has a minimum, $f_{\min}$, and a maximum, $f_{\max}$, so that

$$f_{\min} \leq f(x) \leq f_{\max} \qquad \forall x \in M. \tag{B.2}$$

This means that the whole manifold M can be decomposed in the level sets of f: in fact, one can build M starting from $f^{-1}(f_{\min})$ and then adding continuously to it all the other level surfaces up to $f^{-1}(f_{\max})$. To be more precise, one defines the "part of M below a" as

$$M_a = f^{-1}(-\infty, a] = \{x \in M : f(x) \leq a\}, \tag{B.3}$$

i.e., each M_a is the set of the points $x \in M$ such that the function $f(x)$ does not exceed a given value a; as a is varied between $f_{\min}$ and $f_{\max}$, M_a describes the whole manifold M.

For our purposes, we need to restrict the class of functions we are interested in to the class of *Morse functions*, which are defined as follows. Given a manifold M of dimension n and a smooth function $f: M \mapsto \boldsymbol{R}$, a point $x_c \in M$ is called a *critical point* of f if $\mathrm{d}f(x_c) = 0$, while the value $f(x_c)$ is

called a *critical value*. The function f is called a *Morse function* on M if its critical points are nondegenerate, i.e., if the Hessian matrix of f at x_c, whose elements in local coordinates are

$$H_{ij} = \frac{\partial^2 f}{\partial x^i \partial x^j} \tag{B.4}$$

has rank n, i.e., has only non-zero eigenvalues. This means that there are no directions along which one could move the critical point, so that there are no lines (or surfaces, or hypersurfaces) made of critical points. As a consequence, one can prove that the critical points x_c of a Morse function, and also its critical values, are isolated. It can be proved also that Morse functions are generic: the space of the Morse functions is a *dense* subset of the space of the smooth functions from M to $\boldsymbol{R}$. A level set $f^{-1}(a)$ of f is called a *critical level* if a is a critical value of f, i.e., if there is at least one critical point $x_c \in f^{-1}(a)$.

The main results of Morse theory are the following:

1. If the interval $[a,b]$ contains no critical values of f, then the topology of $f^{-1}[a,v]$ does not change for any $v \in (a,b]$. This result[28] is sometimes called the *non-critical neck theorem*. The reason for this terminology will be made clear in the following.
2. If the interval $[a,b]$ contains critical values, the topology of $f^{-1}[a,v]$ changes in correspondence with the critical values themselves, in a way which is completely determined by the properties of the Hessian of f at the critical points.
3. Some topological invariants of M, i.e., quantities that are the same for all the manifolds which have the same topology as M, so that they characterize unambigously the topology itself, can be estimated and sometimes computed exactly once all the critical points of f are known.

Without giving explicit proofs, which can be found in Ref. [123], let now us discuss in more detail items 1–3 above.

B.1. The non-critical neck theorem

If there are no critical values in the interval $[a,b]$, there exists a diffeomorphism which sends $f^{-1}[a,b]$ into the Cartesian product $f^{-1}(a) \times [a,b]$. This means that the shape of $f^{-1}[a,b]$ is that of a multi-dimensional cylinder, or a neck (from which the name "non-critical neck"), if $f^{-1}(a)$ is simply connected, because the Cartesian product of a circle and an interval is a cylinder. This might be better understood with the aid of a two-dimensional example. Suppose that M is two-dimensional, and that the level set $f^{-1}(a)$ is topologically equivalent to a circle (see Fig. 29).

Then one can construct a diffeomorphism explicitly as the flow of the gradient vector field of f, ∇f, whose flow lines are orthogonal to the level surfaces of f and are depicted as the lines with the

[28] We note that this result is valid even if f is not a Morse function; it is sufficient that it is a smooth function.

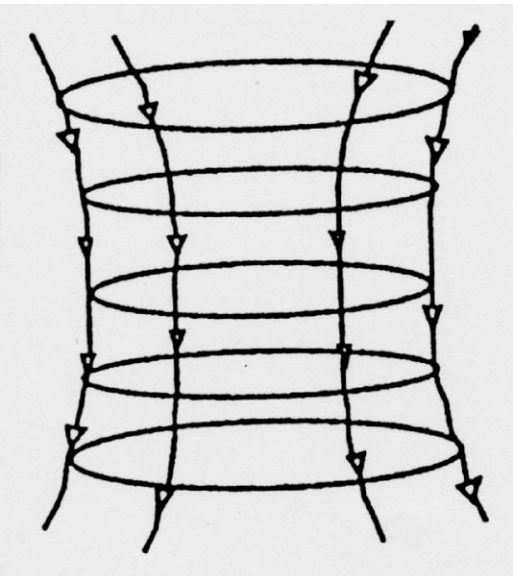

Fig. 29. A non-critical neck. The lines with the arrows are the flow lines of ∇f, and the ellipses are the level sets of f.

arrows in Fig. 29. This flow has no singularities if there are no critical values of f, so that the level set $f^{-1}(a)$ is transported up to $f^{-1}(b)$ along the flow lines of ∇f without changing its topology. As a consequence,

$$f^{-1}[a,b] \approx f^{-1}(a) \times [a,b] \approx f^{-1}(b) \times [a,b], \tag{B.5}$$

where "$x \approx y$" must be read as "x is diffeomorphic to y".

B.2. Critical points and topology changes

In the neighborhood of a regular point P, $N(P)$, there always exists a coordinate system such that f can be written as its first-order Taylor expansion,[29] setting the origin of such coordinates in P, in the form

$$f(x) = f(0) + \frac{\partial f}{\partial x^i} x^i + \cdots \quad \forall x \in N(P). \tag{B.6}$$

Geometrically, this means that in the neighbourhood of a regular point the level sets of f look like hyperplanes in $\boldsymbol{R}^n$, because they are the level sets of a *linear* function.

But what if P is a critical point of f? A fundamental result by Morse, called the *Morse lemma*, is that if f is a Morse function then there always exists in $N(P)$ a coordinate system (called a *Morse*

[29] This follows from the implicit function theorem.

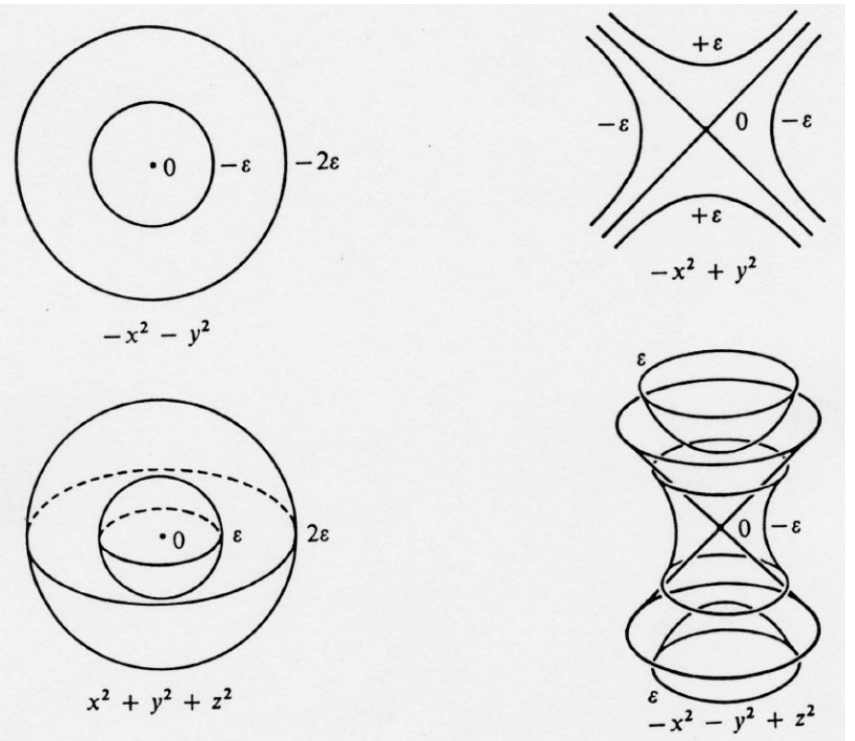

Fig. 30. Some examples of ε-level sets near a critical point (the critical value of the function is set to 0). Upper left: $n = 2$, critical point of index $k = 2$; upper right: $n = 2$, critical point of index $k = 1$; lower left: $n = 3$, critical point of index $k = 0$; lower right: $n = 3$, critical point of index $k = 2$.

chart) such that f is given by its *second-order* Taylor polynomial

$$f(x) = f(0) + \frac{\partial^2 f}{\partial x^i \partial x^j} x^i x^j + \cdots \quad \forall x \in N(P). \tag{B.7}$$

With a suitable rotation of the coordinate frame, $\{x^i\} \mapsto \{y^i\}$, expansion (B.7) can always be reduced to the canonical diagonal form

$$f(y) = f(0) - \sum_{i=1}^{k} (y^i)^2 + \sum_{i=k+1}^{n} (y^i)^2 + \cdots \quad \forall u \in N(P). \tag{B.8}$$

Close to P, the level sets of f are the level sets of a *quadratic* function, so that, geometrically, they are non-degenerate quadrics, like hyperboloids or ellipsoids, which become degenerate at P. The number of negative eigenvalues of the Hessian matrix, k, is called the *index* of the critical point. Passing through the critical level, the shape of the level sets of f changes dramatically, in a way that is completely determined by the index k. Some examples in two and three dimensions are given in Fig. 30.

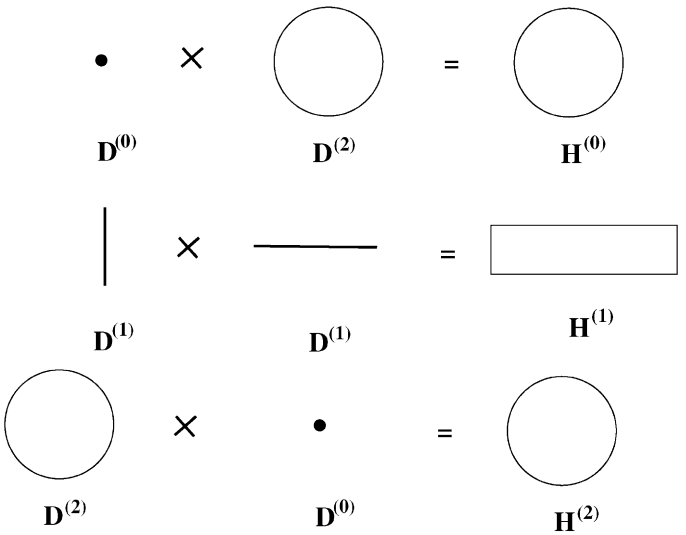

Fig. 31. Two-dimensional handles: $H^{(0)}$ is the product of a 0-disk (a point) and a 2-disk, so that it is a 2-disk; $H^{(1)}$ is the product of two 1-disks, i.e., of two intervals, so that it is a strip; $H^{(2)}$ is again a 2-disk as $H^{(0)}$ is.

The change undergone by the submanifolds M_a as a critical level is passed is described using the concept of "attaching handles". A k-handle $H^{(k)}$ in n dimensions ($0 \leq k \leq n$) is a product of two disks, one k-dimensional (D^k) and the other $(n-k)$-dimensional (D^{n-k}):

$$H^{(k)} = D^k \times D^{n-k}. \tag{B.9}$$

In two dimensions, we can have either 0-handles, which are two-dimensional disks or 1-handles, which are the product of two one-dimensional disks, i.e., of two intervals, so that they are stripes, or 2-handles, which are again two-dimensional disks (Fig. 31). In three dimensions (see Fig. 32), we have 0-handles which are solid spheres, 1-handles which are the product of a disk and an interval, so that they are solid cylinders, 2-handles which are the same as 1-handles, and 3-handles which are the same as 0-handles (Fig. 31).

In more than three dimensions it is difficult to visualize handles: however, there is still the duality of the $n = 2$ and 3 cases, i.e., k and $n - k$ handles are topologically equivalent.

Having defined handles, we can state the main result of Morse theory as follows.

Theorem. *Suppose that there is one* (*and only one*) *critical value* c *in the interval* $[a, b]$, *and that it corresponds to only one critical point of index* k. *Then the manifold* M_b *arises from* M_a *by attaching a* k*-handle, and the transition occurs precisely at the critical level* c. *Everything goes in the same way if there are* $m > 1$ *critical points, with indices* $k_1, \ldots, k_m$ *on the critical level* $f^{-1}(c)$; *in this case* M_b *arises from* M_a *by attaching* m *disjoint handles of types* $k_1, \ldots, k_m$.

Let us see how this works in a simple example. Consider as our manifold M a two-dimensional torus standing on a plane (think of a tyre in a ready-to-roll position), and define a function f on it as

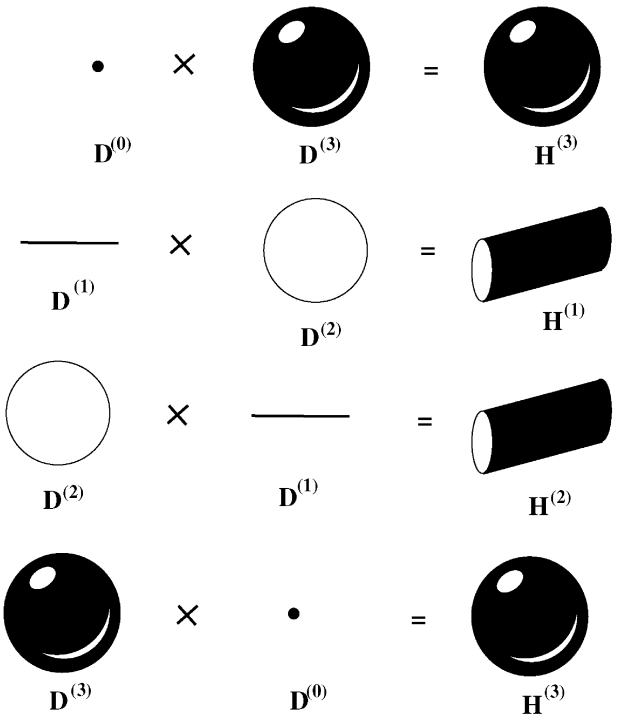

Fig. 32. Three-dimensional handles: $H^{(0)}$ is the product of a 0-disk (a point) and a 3-disk (a ball), so that it is a ball; $H^{(1)}$ is the product of a 1-disk (an interval) and a 2-disk, so that it is a tube; $H^{(2)}$ is as $H^{(1)}$, and $H^{(3)}$ is as $H^{(0)}$.

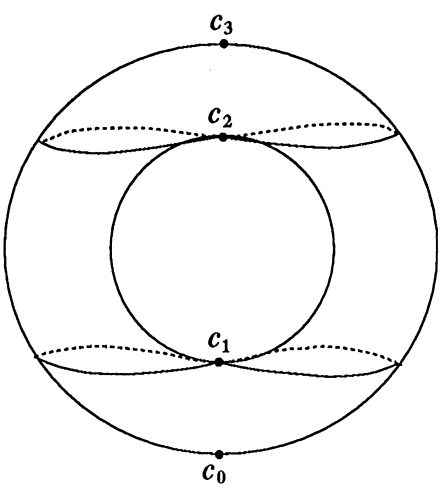

Fig. 33. The critical points and critical levels of the height function on a two-dimensional torus.

the height of a point of M above the floor level. If the z-axis is vertical, f is the orthogonal projection of M onto the z-axis. Such a function has four critical points, and the corresponding four critical levels of f, which will be denoted as c_0, c_1, c_2, c_3, respectively, are depicted in Fig. 33. We can build

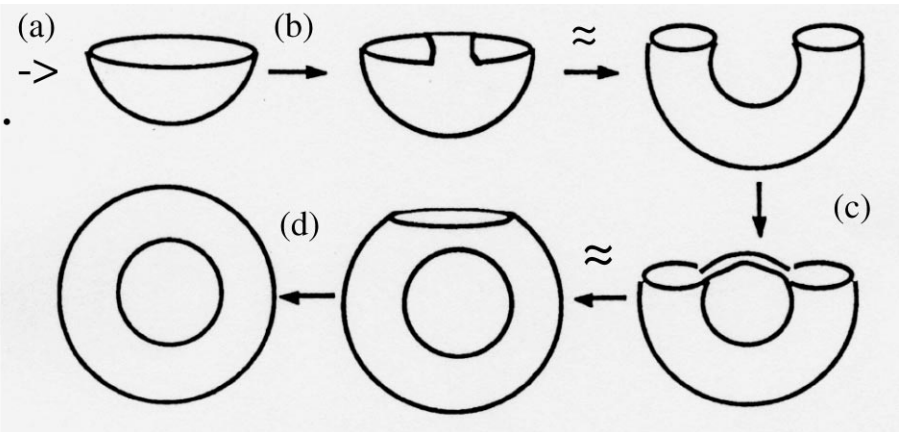

Fig. 34. The building of a two-dimensional torus by attaching handles: (a) attaching a 0-handle, (b) attaching a 1-handle, (c) attaching a 1-handle again, (d) attaching a 2-handle to complete the torus. The symbol " $\approx$ " means "is diffeomorphic to".

our torus in separate steps: each step will correspond to the crossing of a critical level of f. The steps are pictorially described in Fig. 34. As long as $a < 0$, the manifold M_a is empty. At $a = c_0 = 0$ we cross the first critical value, corresponding to a critical point of index 0. This means that we have to attach a 0-handle (a disk) to the empty set. Any M_a with $0 < a < c_1$ is diffeomorphic to a disk, as we can see by cutting a torus at any height between 0 and c_1 and throwing away the upper part. At c_2 we meet the second critical point, which now has index 1, so that we have to attach a 1-handle (a stripe) to the previous disk, obtaining a sort of a basket. Such a basket can be smoothly deformed into a U-shaped tube: in fact, if we cut a torus at any height between c_1 and c_2 and we throw away the upper part, we get a U-shaped tube. The third critical point c_2 is again a point of index 1, so we have to glue another stripe to the tube. What we obtain can be smoothly deformed into a full torus with only the polar cap cut away from it. The last critical point has index 2, so that the crossing of it corresponds to the gluing of a 2-handle (a disk), which is just the polar cap we needed to complete the torus.

B.3. Topological invariants

Morse theory can be used also to give estimates, and sometimes to compute exactly, some topological invariants of our manifold M. For a (two-dimensional) surface, a very important topological invariant is the *genus* g, which equals the number of handles of the surface. The generalization to n dimensions of the genus is given by the set of the *Betti numbers* $b_k(M)$, $0 \leq k \leq n$, which are the fundamental topological invariants of an n-dimensional manifold, and completely describe its topology. For hyperspheres, all the Betti numbers are zero. Using the Betti numbers we can obtain another topological invariant, the *Euler characteristic* $\chi(M)$, which is nothing but the alternating sum of the b_k:

$$\chi(M) \equiv \sum_{k=0}^{n} (-1)^k b_k(M). \tag{B.10}$$

For (two-dimensional) surfaces, $\chi = 2 - 2g$ holds.

Now let us consider a generic Morse function f on M and let us define the *Morse numbers* $\mu_k(M)$ as follows: μ_k is the total number of critical points of f on M which have index k. It turns out that the Morse numbers of a manifold are upper bounds of the Betti numbers, i.e., the following (*weak*) *Morse inequalities* hold:

$$b_k(M) \leq \mu_k(M), \quad k = 0, \ldots, n. \tag{B.11}$$

Actually, a result stronger than Eqs. (B.11) holds, which states that alternate sums of two, three, four, … subsequent Betti numbers are bounded from above by the alternate sums of the corresponding Morse numbers. Starting from this result one can prove the following identity:

$$\chi(M) = \sum_{k=0}^{n} (-1)^k \mu_k(M) \tag{B.12}$$

and this provides a way of computing exactly the Euler characterstic of a manifold once all the critical points of a Morse function are known.

Among all the Morse functions on a manifold M, there is a special class (called *perfect* Morse functions) for which the Morse inequalities (B.11) hold as equalities. Perfect Morse functions characterize completely the topology of a manifold. It is possible to prove that the height function on the torus we considered above is a perfect Morse function [123]. However, there are no simple general recipes to construct perfect Morse functions (this is actually an active area of research).

Appendix C. Chaos in Hamiltonian dynamical systems

For a long time the equations of Newtonian mechanics have been the paradigm of classical determinism. Only quite recently has it been realized that "determinism" and "predictability" are far from being the same concept, and that predictability also requires the *stability* of the solutions of the dynamical differential equations. Determinism implies that, once an initial condition is given, the trajectory is uniquely determined for all later times; stability means that two initially close trajectories will remain close in the future (more precisely, their distance will grow slower than a power of the time). If this is not true, it becomes impossible to predict the evolution of a system even for very small times, as explained, for instance, by Feynman [124]:

> It is true classically that if we knew the position and the velocity of every particle in the world, or in a box of gas, we could predict exactly what would happen. And therefore the classical world is deterministic. Suppose, however, that we have a finite accuracy and do not know *exactly* where just one atom is, say to one part in a billion. Then as it goes along it hits another atom, and because we did not know the position better than one part in a billion, we find an even larger error in the position after the collision. And that is amplified, of course, in the next collision, so that if we start with only a tiny error it rapidly magnifies to a very great uncertainty. To give an example: if water falls over a dam, it splashes. If we stand nearby, every now and then a drop will land on our nose. This appears to be completely random, yet such a behavior would be predicted by purely classical laws. The exact position of all the drops

depends upon the precise wigglings of the water before it goes over the dam. How? the tiniest irregularities are magnified in falling, so that we get complete randomness. Obviously, we cannot predict the position of the drops unless we know the motion of the water *absolutely exactly*.

Speaking more precisely, given an arbitrary accuracy, no matter how precise, one can find a time long enough that we cannot make predictions valid for that long a time. Now the pint is that this length of time is not very large. It is not that the time is millions of years if the accuracy is one part in a billion. The time goes, in fact, only logarithimically with the error, and it turns out that in only a very, very tiny time we lose all our information. If the accuracy is taken to be one part in billions and billions and billions – no matter how many billions we wish, provided we do stop somewhere – then we can find a time less than the time it took to state the accuracy – after which we can no longer predict what is going to happen!

As long as non-linear dynamical systems are considered, stability is the exception rather than the rule. Even if this relies – at least from a conceptual point of view – upon mathematical results which have been known since the end of the last century, its importance has only been completely realized with the aid of a new and powerful approach: numerical simulation. The very complicated structure of some trajectories which can arise in non-linear dynamical systems was discovered by Poincaré [1] in the late 19th century, but the physics community became fully aware of the existence and of the meaning of these structures only they were visualized by computer simulation in the work of Hénon and Heiles [61].

The instability we are referring to is known as intrinsic stochasticity of the dynamics, or "deterministic chaos". These terms mean that the dynamics, being completely deterministic, yet exhibits some features that make it indistinguishable from a random process. The characteristic feature of a chaotic system, which is at the basis of the unpredictability of its dynamics, is the sensitive (exponential) dependence on initial conditions: the distance between two trajectories which originate in very close-by points in phase space grows exponentially in time so that the system looses the memory of its initial conditions. Regular dynamics, i.e., quasiperiodic motion, is – as far as conservative systems are considered – a "weak" property, because it is destroyed by very small perturbations of the system. On the contrary, chaos is a strong property, because given a dynamical system where chaos is present, in many cases it will be present even after the system has been subjected to significant perturbations [125].

Here we recall briefly the main concepts of the theory of Hamiltonian dynamical systems which are necessary for the understanding of the material on chaos presented in this report. The main goal of this Appendix is then to provide the reader with a definition of the Lyapunov exponents and of a motivation for the introduction of these quantities as a "measure" of chaos in a dynamical system.

A very good introduction to the subject is given in Lichtenberg and Lieberman's classic book [125], and, at a more pedagogical level, in Tabor's [126] and Ott's textbooks [127]. An interesting selection of reprints can be found in MacKay and Meiss [128]. We assume the reader is familiar, at least at a basic level, with the concepts of *ergodicity* and *mixing*. A discussion on these topics would be far beyond the scope of the present Report; a good introduction can be found in any of the references just mentioned above.

C.1. A simple example of chaotic dynamics: the perturbed pendulum

Throughout the paper we have been concerned with Hamiltonian dynamical systems with a large number of degrees of freedom. However, the main features of chaos can be better appreciated starting with an example of a system with only one degree of freedom, subjected to an external perturbation: the forced pendulum. Although the behaviour of many degree of freedom is much more complicated, nevertheless some of the essential features of chaos are already present in this simple example.

The forced pendulum is a system obeying the following equation of motion:

$$\ddot{q} + \sin q = \varepsilon \sin(q - \omega t). \tag{C.1}$$

The phase space of the forced pendulum is three dimensional, because in addition to the coordinate q and to the momentum $p = \dot{q}$, one has to take into account also the time t, because the forcing term on the right-hand side of Eq. (C.1) depends explicitly on time. The forced pendulum is, however, a Hamiltonian system, and since the dimension of the phase space of autonomous Hamiltonian dynamical systems is $2N$, where N is the number of degrees of freedom, it is customary to refer to systems like the forced pendulum as to systems with "1.5 degrees of freedom".

As long as $\varepsilon = 0$, the system obeying Eq. (C.1) is a simple pendulum, and its Hamiltonian

$$H = \frac{p^2}{2} - \cos q \tag{C.2}$$

is an integral of motion, so that its value, the energy E, is a constant of the motion and the system is *integrable*, as every one-degree-of-freedom autonomous Hamiltonian systems is. The word "integrable" is used here in a wider sense than its immediate meaning "such that the equations of motion can be solved"; a Hamiltonian system is integrable when it has a sufficiently large number of integrals of motion (N, for an autonomous system with N degrees of freedom), such that its trajectories do not explore the whole phase space, but are confined to lower-dimensional subsets called *invariant tori*.[30] When $N = 1$, each invariant torus coincides with a trajectory. Some of these are depicted in Fig. 35. We remark that there are two distinct classes of trajectories: oscillations, which correspond to bounded motions, and rotations, which are unbounded. The two classes are separated by a curve called the *separatrix*. The separatrix is the trajectory pursued by the pendulum when it starts precisely at the unstable equilibrium point $(p, q) = (0, \pm \pi)$ with $E = 1$, i.e., just the energy that is required to come back to the same point (note that $q = \pi$ and $-\pi$ must be identified). The motion on the separatrix requires an infinite amount of time.

But what happens if $\varepsilon \neq 0$? Once $\varepsilon \neq 0$, no matter how small, system (C.1) is no longer integrable, and the separatrix, which was a *unique* curve in the $\varepsilon = 0$ case, splits into *two* distinct invariant curves. These curves must intersect transversally each other infinitely many times, as Poincaré showed for the first time [1]. These intersections are referred to as *homoclinic intersections*, and force the trajectories to fold themselves giving rise to a very complicated structure: in Poincaré's own words [1]

[30] The origin of the term is as follows: such subsets are *invariant* because if a trajectory starts on one of them, it remains there forever; they are called *tori*, because they are topologically equivalent to multi-dimensional tori.

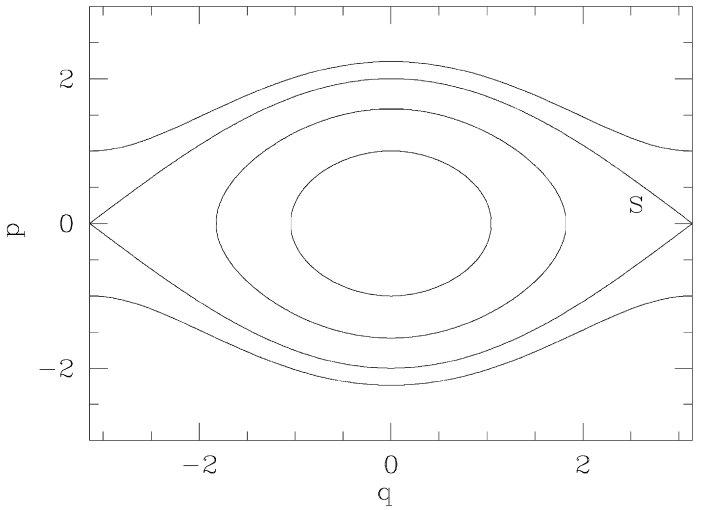

Fig. 35. Phase-space trajectories of a simple pendulum. The closed curves are the oscillations, the curves above and below the separatrix S are clockwise and counterclockwise rotations, respectively.

these intersections form a sort of texture, or of a net whose meshes are infinitely tight; each of these two curves can never intersect itself, but has to fold in a complicated way as to intersect all the meshes of the net an infinite number of times. One is amazed by the complexity of this picture, which I do not even attempt to draw.

As a consequence of the presence of these intersections, in a neighborhood of the region in phase space which was occupied by the separatrix in the integrable case, a so-called *chaotic sea* suddenly appears. The chaotic sea is the region irregularly filled by dots in Fig. 36, where a two-dimensional section[31] of the 3-d phase space of the system is shown. If we now follow the evolution of two intially close points in the chaotic sea, we see that their separation grows *exponentially* in time, so that the dynamics in the chaotic sea is unpredictable.

The appearance, in phase space, of irregular regions like the chaotic sea could justify by itself the use of the term "chaotic dynamics". However, there are also other properties of the dynamics described by Eq. (C.1) which justify the use of such a term. For example, if we introduce a symbolic coding of the dynamics in which the symbol 0 is associated with each passage through the point $q = 0$ with $\dot{q} > 0$ and the symbol 1 to each passage through the same point with $\dot{q} < 0$, then given any bi-infinite sequence of zeros and ones, for example generated by coin tosses, this sequence corresponds to a real trajectory of system (C.1). Aspects of the motion of the system, though deterministic, is thus indistinguishable from a random process.

We can intuitively understand the origin of such a behaviour if we think that when the phase point is on a trajectory very close to the separatrix, an infinitesimal variation may *qualitatively*

[31] This section has been obtained as a stroboscopic Poincaré section [126], so that each point on the plot corresponds to an intersection of a trajectory of the system with the planes $t = 2n\pi/\omega$.

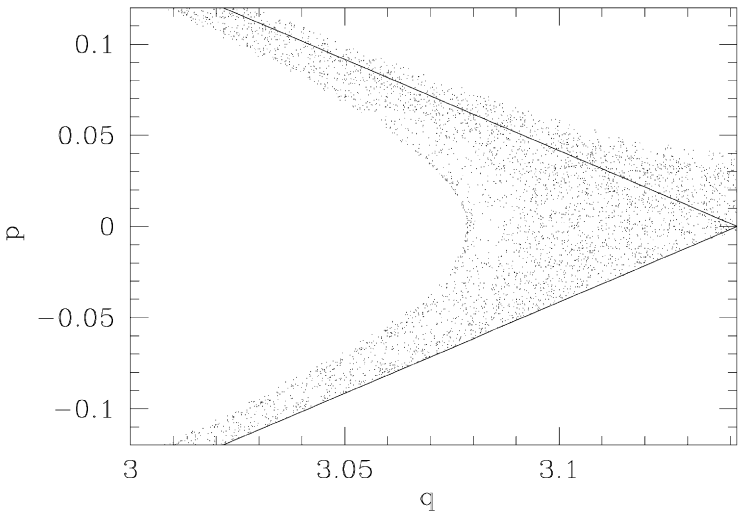

Fig. 36. Section of the phase space of a perturbed pendulum, showing the appearance of chaotic seas close to the separatrix of the unperturbed system (the solid line in the figure). The dots are obtained from a single trajectory issuing from a point very close to the unperturbed separatrix. The amplitude of the perturbation is $\varepsilon = 10^{-4}$.

change the character of the motion (e.g., from oscillations to rotations). This is an example of the sensitive dependence on initial conditions that, in general, leads to the exponential separation of initially close-by trajectories.

This example is extremely simple but contains the essential features of the problem. In fact, even the case of a Hamiltonian system with N degrees of freedom can be treated in a similar way and shows analogous, though much more complicated, results. In that case, there is no need of an external forcing to get chaos, for as $N > 1$ an autonomous, non-linear Hamiltonian system is generically non-integrable (the integrable systems being a very small subset of all the possible systems). However, even in the simple low-dimensional cases, by means of concepts like homoclinic intersections, it is possible only to give a qualititative description of the onset of chaos, but a quantitative description of the stochastic regions is impossible, i.e., it is impossible to compute *how fast* two initially close-by points will separate. For N-dimensional systems, the situation is obviously even worse: there exists a method, which is a generalization of Poincaré's method, obtained by Mel'nikov [129] and later by Arnol'd [130], which allows one to show the existence of homoclinic intersections near separatrices for very small perturbations even for large systems, but again there is no possibility of describing quantitatively the stochastic regions.

To obtain quantitative informations on chaotic dynamics we must introduce the concept of Lyapunov exponent.

C.2. Lyapunov exponents

We now give a definition and an explanation of the Lyapunov exponents. Our discussion will be aimed at showing how to define and compute the Lyapunov exponents for a dynamical system which is defined by a system of ordinary differential equations, i.e., for a flow, because Hamiltonian

dynamical systems belong to this class. For a more general discussion of Lyapunov exponents, see [131].

From a physical point of view, given a trajectory of a dynamical system, it is important to find answers to the following questions: Is the trajectory chaotic? And if so, how strong is the chaos, or how fast do two initially close-by points separate in phase space, i.e., how long should we wait until the system exhibits its chaotic feature? The concept of Lyapunov exponents is introduced to answer these questions, since Lyapunov exponents are defined in order to provide an average measure of the rate of exponential divergence of nearby orbits in phase space, which is the distinctive feature of chaos.

Lyapunov exponents are defined for a given trajectory of a dynamical system, and this allows us to give a definition of a chaotic trajectory as follows: a trajectory is said to be *chaotic* if its (largest) Lyapunov exponent is *positive*.

We now give a definition of the Lyapunov exponents. Let us consider a dynamical system whose trajectories in an n-dimensional phase space M are the solutions of the following system of ordinary differential equations:

$$\begin{aligned} \dot{x}_1 &= X_1(x_1,\dots,x_n), \\ &\vdots \qquad\qquad \vdots \\ \dot{x}_n &= X_n(x_1,\dots,x_n). \end{aligned} \tag{C.3}$$

If we denote by $x(t) = (x_1(t),\dots,x_n(t))$ a given trajectory whose initial condition is $x(0)$, and by $y(t)$ another trajectory which is initially close to $x(t)$, and we denote by $\xi(t)$ the vector

$$\xi(t) = y(t) - x(t), \tag{C.4}$$

then the evolution of ξ describes the separation of the two trajectories in phase space. The vector ξ is assumed to obey the linearized equations of motion, because it is assumed to be initially small. These equations are, as can be shown by inserting Eq. (C.4) into the equations of motion (C.3) and expanding in a power series up to the linear terms

$$\begin{aligned} \dot{\xi}_1 &= \sum_{j=1}^{n} \left(\frac{\partial X_1}{\partial x_j}\right)_{x(t)} \xi_j, \\ &\vdots \qquad\qquad \vdots \\ \dot{\xi}_n &= \sum_{j=1}^{n} \left(\frac{\partial X_n}{\partial x_j}\right)_{x(t)} \xi_j \end{aligned} \tag{C.5}$$

and are referred to as the *tangent dynamics equations*,[32] which we already wrote in the main text in the particular case of a standard Hamiltonian system (see Eq. (25)). Note that (C.5) is a system of linear differential equations, whose coefficients, however, depend on time. According to definition (C.4), the norm $|\xi|$ of the vector ξ, i.e.,

$$|\xi(t)| = \left[\sum_{i=1}^{n} \xi_i^2(t)\right]^{1/2} \tag{C.6}$$

[32] This notation follows from that the dynamics of the vector ξ takes place in the tangent space $T_{x(t)}M$ of the phase space M.

measures the distance of the two trajectories as a function of t. If the trajectory $x(t)$ is unstable, all its perturbations grow exponentially, so that $|\xi(t)| \propto \exp(\lambda t)$. If the elements of the Jacobian matrix $\partial X_i/\partial x_j$, which are the coefficients of the linear equations (C.5), were either constant or periodic, it would be possible to solve the system, but, since the Jacobian matrix depends on the trajectory $x(t)$, its entries are in general neither constant nor periodic, so that the rate of exponential divergence varies with time. Therefore, one introduces an asymptotic rate of exponential growth of ξ as the *Lyapunov exponent* λ

$$\lambda = \lim_{t\to\infty} \frac{1}{t} \log \frac{|\xi(t)|}{|\xi(0)|}, \tag{C.7}$$

which measures the degree of instability of a trajectory: if λ is positive, the trajectory is unstable with a characteristic time λ^{-1}. In principle, λ depends on both the initial values of x, $x(0)$, and of ξ, $\xi(0)$. However, Oseledeč [132] has shown that limit (C.7) exists, is finite and can assume only one of the n values

$$\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n . \tag{C.8}$$

The set $\{\lambda_i\}$ is the called *Lyapunov spectrum*. The exponent λ defined in (C.7) takes the n different values of the spectrum as the initial condition ξ in the tangent space $T_{x(0)}M$ is varied; the latter admits a decomposition in linear subspaces

$$T_{x(0)}M = E_1 \oplus E_2 \oplus \cdots \oplus E_n \tag{C.9}$$

and each λ_i is associated with the corresponding subspace E_i, in that a vector $\xi(0) \in E_i$ will exponentially grow with the exponent λ_i. If there exists on the phase space M a probability measure μ, which is ergodic and invariant for the dynamics on M, then the numbers λ_i do *not* depend on the initial condition $x(0)$, apart from a possible subset of initial conditions of measure zero with respect to μ.

In practice, the evolution of the norm of a tangent vector is sensitive only to the first – the largest – exponent, because a generic initial vector $\xi(0)$ will have a nonvanishing component in the E_1 subspace, so that the largest exponent λ_1 will always dominate in the long-time limit: choosing ξ at random with respect to a uniform distribution we have $\lambda = \lambda_1$ with probability one. This means that Eq. (C.7) provides a practical definition for the *largest* Lyapunov exponent λ_1, which we have always denoted simply by λ in the main text.

Let us now apply the above to a standard Hamiltonian system, whose Hamiltonian is of the form (1); the dimension of the phase space is $n = 2N$, and the equations of motion (C.3) are now Hamilton's equations

$$\begin{aligned}
\dot{q}_1 &= p_1 , \\
&\vdots \\
\dot{q}_N &= p_N , \\
\dot{p}_1 &= -\frac{\partial V}{\partial q_1} , \\
&\vdots \\
\dot{p}_N &= -\frac{\partial V}{\partial q_N}
\end{aligned} \tag{C.10}$$

and also the linearized dynamics (C.5) can be cast in the canonical form

$$\begin{aligned}
\dot{\xi}_1 &= \xi_{N+1}\,,\\
&\vdots \qquad\qquad \vdots\\
\dot{\xi}_N &= \xi_{2N}\,,\\
\dot{\xi}_{N+1} &= -\sum_{j=1}^{N}\left(\frac{\partial^2 V}{\partial q_1 \partial q_j}\right)_{q(t)}\xi_j\,,\\
&\vdots \qquad\qquad \vdots\\
\dot{\xi}_{2N} &= -\sum_{j=1}^{N}\left(\frac{\partial^2 V}{\partial q_N \partial q_j}\right)_{q(t)}\xi_j\,.
\end{aligned} \tag{C.11}$$

This equation was already introduced as Eq. (25) in Section 3, and is usually referred to as the tangent dynamics equation for Hamiltonian systems. To measure the largest Lyapunov exponent λ in a numerical simulation, one integrates numerically both Eqs. (C.10) and (C.11), and then makes use of definition (C.7), which can be rewritten, in this case, as

$$\lambda = \lim_{t\to\infty}\frac{1}{t}\log\frac{[\xi_1^2(t)+\cdots+\xi_N^2(t)+\dot{\xi}_1^2(t)+\cdots+\dot{\xi}_N^2(t)]^{1/2}}{[\xi_1^2(0)+\cdots+\xi_N^2(0)+\dot{\xi}_1^2(0)+\cdots+\dot{\xi}_N^2(0)]^{1/2}}\,, \tag{C.12}$$

where we have used that $\dot{\xi}_i = \xi_{i+N}$ (see Eq. (C.11)). More precisely, in a numerical simulation one uses the discretized version of Eq. (C.12), i.e.,

$$\lambda = \lim_{m\to\infty}\frac{1}{m}\sum_{i=1}^{m}\frac{1}{\Delta t}\log\frac{|\xi(i\Delta t+\Delta t)|}{|\xi(i\Delta t)|}\,, \tag{C.13}$$

where after a given number of time steps Δt, the value of $|\xi|$ has to be renormalized to a fixed value, in order to avoid overflow [133].

Definition (C.7) does not allow one to measure the other exponents of the Lyapunov spectrum. To measure them, one has to observe that they can be related to the growth of *volumes* in the tangent space. A two-dimensional area V_2 in the tangent space, spanned by two linearly independent tangent vectors $\xi^{(1)}$ and $\xi^{(2)}$, will expand according to

$$V_2(t) \propto \exp[(\lambda_1+\lambda_2)t]\,, \tag{C.14}$$

a three-dimensional volume, as

$$V_3(t) \propto \exp[(\lambda_1+\lambda_2+\lambda_3)t]\,, \tag{C.15}$$

and so on, so that, choosing $k \le n$ linearly independent and normalized vectors $\xi^{(1)}, \xi^{(2)}, \ldots, \xi^{(k)} \in T_x M$ we obtain

$$\lim_{t\to\infty}\frac{1}{t}\log|\xi^{(1)}(t)\wedge\xi^{(2)}(t)\wedge\cdots\wedge\xi^{(k)}(t)| = \sum_{i=1}^{k}\lambda_i\,. \tag{C.16}$$

Therefore algorithm (C.13) can be generalized to obtain an algorithm to compute numerically the whole Lyapunov spectrum [133]. However, such a computation is very hard when the number N is large.

The sum of *all* the n Lyapunov exponents in the Lyapunov spectrum, $\sum_{i=1}^{n}\lambda_i$, measures the expansion rate of n-volumes in phase space. Therefore, for a Hamiltonian system

$$\sum_{i=1}^{2N}\lambda_i = 0\,, \tag{C.17}$$

because volumes in phase space are conserved. In addition, for Hamiltonian systems a result stronger than (C.17) holds, i.e., there is a symmetry in the Lyapunov spectrum such that

$$\lambda_i = -\lambda_{2N-i+1}\,. \tag{C.18}$$

Eq. (C.18) for Hamiltonian systems is a consequence of the symplectic structure of the Hamilton's equations [134], however it has been recently generalized to a class of non-Hamiltonian systems [135].

The numerical integration of the Eqs. (C.5) and the consequent measure of λ – or of the spectrum $\{\lambda_i\}$ when it is possible in practice – is the standard technique to characterize Hamiltonian chaotic dynamics. An operative definition of a chaotic dynamical system can be stated as follows: a system is chaotic if it has at least one positive and one negative Lyapunov exponent. In fact, this ensures that the system shows (almost everywhere with respect to the ergodic measure μ used to define the Lyapunov exponents) the distinctive features of chaos as described in the example of the forced pendulum. In fact, the presence of a positive exponent ensures the presence of a exponential divergence of nearby orbits, and the presence of a negative one ensures that they also fold and mix in a very complicated way, so that they can produce those structures we refrred to as "chaotic seas". However, as long as autonomous Hamiltonian systems are considered, the anti-symmetry of spectrum (C.18) ensures that the presence of a positive exponent implies the presence of a negative one with the same absolute value, so that a single (the largest) positive exponent is sufficient to have chaos; on the contrary, if the largest exponent vanishes the dynamics will be regular. These facts, together with that the largest Lyapunov exponent λ measures the smallest instability time scale, show how natural the use of the value of λ is to measure chaos in such systems.

It is important to specify with respect to what invariant ergodic measure μ the Lyapunov exponents are defined: this may be also a δ-measure concentrated on a single trajectory, in which case we could speak of a chaotic trajectory rather than of a chaotic system. In Hamiltonian systems with a large number of degrees of freedom we expect the microcanonical measure of the chaotic regions to be overwhelmingly larger than the measure of the regular regions; the existence of these regular regions is ensured – at least as long as the system is not too far from an integrable limit – by the Kol'mogorov–Arnol'd–Moser (KAM) theorem [57]. However, from a practical point of view the measure relevant for the definition of the Lyapunov exponent is indeed the microcanonical one. Numerical experiments are in agreement with this expectation for large systems, since no relevant dependence of the Lyapunov exponent on the initial conditions has been detected, and this is the reason why in the main text we have never referred explicitly to any dependence of λ on μ, treating the Lyapunov exponent as any other "thermodynamic" observable. Nevertheless, for small systems (especially $N = 2$ which is the best known case) the simulations show that the measure of the

chaotic regions may be very small in a very large energy range, so that in that case one cannot speak of a truly chaotic system but only of a system in which chaotic and regular regions are simultaneously present (these systems are often referred to as *mixed* systems, as they are in between completely chaotic and regular ones).

Since we are interested in large systems, up to the thermodynamic limit, a number of questions naturally arises: what is the behaviour of the Lyapunov exponents as n increases; does a thermodynamic limit exist for the Lyapunov spectrum, etc. Numerical results [136] have shown that as $n \to \infty$ the Lyapunov spectrum $\{\lambda_i\}$ appears indeed to converge to a well-behaved function

$$\lambda(x) = \lim_{n\to\infty} \lambda_{xn}\,. \tag{C.19}$$

The function $\lambda(x)$ is a non-increasing function of $x \in [0,1]$. Some rigorous work in this respect has been recently done by Sinai [137]. The existence of a limiting Lyapunov spectrum in the thermodynamic limit has many important consequences that we will not review here; a good discussion can be found in Ref. [134]. We only want to remark here that the existence of a thermodynamic limit for the Lyapunov spectrum implies that the largest Lyapunov exponent is expected to behave as an intensive quantity as N increases.

Appendix D. The stochastic oscillator equation

In the following, we will briefly describe how to cope with the stochastic oscillator problem which we encountered in Section 4.1.3. The discussion closely follows Van Kampen [67] where all the details can be found.

A stochastic differential equation can be put in the general form

$$F(x, \dot{x}, \ddot{x}, \ldots, \Omega) = 0\,, \tag{D.1}$$

where F is an assigned function and the variable Ω is a random process, defined by a mean, a standard deviation and an autocorrelation function. A function $\xi(\Omega)$ is a solution of this equation, if $F(\xi(\Omega), \Omega) = 0\ \forall\Omega$. If Eq. (D.1) is linear of order n, it is written as

$$\dot{\boldsymbol{u}} = \boldsymbol{A}(t, \Omega)\boldsymbol{u}\,, \tag{D.2}$$

where

$$\boldsymbol{u} = \begin{pmatrix} u_1 \\ u_2 \\ u_3 \\ \vdots \\ u_n \end{pmatrix} = \begin{pmatrix} x \\ \dot{x} \\ \ddot{x} \\ \vdots \\ x^{(n)} \end{pmatrix} \tag{D.3}$$

and $\boldsymbol{A}$ is an $n \times n$ matrix whose elements $A_{\mu\nu}(t)$ depend randomly on time.

For the purposes of our work, we are interested in the evolution of the quantities $u_\nu u_\mu$, rather than of the u_μ's themselves. The products $u_\mu u_\nu$'s obey the differential equation

$$\frac{\mathrm{d}}{\mathrm{d}t}(u_\nu u_\mu) = \sum_{k,\lambda} \tilde{A}_{\nu\mu,k\lambda}(t)(u_k u_\lambda), \tag{D.4}$$

where

$$\tilde{A}_{\nu\mu,k\lambda} = A_{\nu k}\delta_{\mu\lambda} + \delta_{\nu k}A_{\mu\lambda}. \tag{D.5}$$

However, both Eqs. (D.2) and (D.4) have exactly the same form and can be solved using the same procedure, so that we will first illustrate such a procedure in general. Therefore in the following formulae, $\boldsymbol{u}$ refers to a vector whose components are either the u_μ's or the $u_\mu u_\nu$'s, and $\boldsymbol{A}$ denotes either the matrix $\boldsymbol{A}$ in Eq. (D.2) or the matrix $\tilde{\boldsymbol{A}}$ whose elements are given by Eq. (D.5), respectively. Then, we will apply this procedure to the case of the stochastic harmonic oscillator.

Now, solving a linear stochastic differential equation means determining the evolution of the average of $\boldsymbol{u}(t)$, $\langle \boldsymbol{u}(t)\rangle$, where the average is carried over all the realizations of the process. Let us consider the matrix $\boldsymbol{A}$ as the sum

$$\boldsymbol{A}(t,\Omega) = \boldsymbol{A}_0(t) + \alpha\boldsymbol{A}_1(t,\Omega), \tag{D.6}$$

where the first term is Ω-independent and the second one is randomly fluctuating with zero mean. Let us also assume that $\boldsymbol{A}_0$ is time-independent. If the parameter α – which determines the fluctuation amplitude – is small we can treat Eq. (D.2) by means of a perturbation expansion. It is convenient to use the interaction representation, so that we put

$$\boldsymbol{u}(t) = \exp(\boldsymbol{A}_0 t)\boldsymbol{v}(t) \tag{D.7}$$

and

$$\boldsymbol{A}_1(t) = \exp(\boldsymbol{A}_0 t)\boldsymbol{v}(t)\exp(-\boldsymbol{A}_0 t). \tag{D.8}$$

Formally, one is then led to a Dyson expansion for the solution $\boldsymbol{v}(t)$. Then, going back to the previous variables and averaging, the second-order approximation gives

$$\frac{\mathrm{d}}{\mathrm{d}t}\langle \boldsymbol{u}(t)\rangle = \left\{\boldsymbol{A}_0 + \alpha^2\int_0^{+\infty} \langle \boldsymbol{A}_1(t)\exp(\boldsymbol{A}_0\tau)\boldsymbol{A}_1(t-\tau)\rangle \exp(-\boldsymbol{A}_0\tau)\,\mathrm{d}\tau\right\}\langle \boldsymbol{u}(t)\rangle. \tag{D.9}$$

Let us remark that, if the stochastic process Ω is Gaussian, Eq. (D.9) is more than a second order approximation: it is exact. In fact, the Dyson series can be written in compact form as

$$\langle \boldsymbol{u}(t)\rangle = T\left[\left\langle \exp\left(\int_0^t \boldsymbol{A}(t')\,\mathrm{d}t'\right)\right\rangle\right]\langle \boldsymbol{u}(0)\rangle, \tag{D.10}$$

where $T[\cdots]$ stands for a time-ordered product. According to Wick's procedure we can rewrite Eq. (D.10) as a cumulant expansion, and when the cumulants of higher than the second order vanish (as in the case of a Gaussian process) one can easily show that Eq. (D.9) is exact.

We now apply this general approach to the case of interest for the main text, i.e., to the stochastic harmonic oscillator equation, which is the the second-order linear stochastic differential equation given by

$$\ddot{x} + \Omega(t)x = 0, \tag{D.11}$$

where $\Omega(t)$ is the random squared frequency, $\Omega = \Omega_0 + \sigma_\Omega \eta(t)$, where Ω_0 is the mean of $\Omega(t)$, σ_Ω is the amplitude of the fluctuations, and $\eta(t)$ is a stochastic process with zero mean. In this case, Eq. (D.2) has the form

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{pmatrix} x \\ \dot{x} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -\Omega & 0 \end{pmatrix}\begin{pmatrix} x \\ \dot{x} \end{pmatrix}. \tag{D.12}$$

In particular, we are interested in obtaining the averaged equation of motion for the second moments. Using Eqs. (D.5) and (D.12), one finds that Eq. (D.4) becomes

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{pmatrix} x^2 \\ \dot{x}^2 \\ x\dot{x} \end{pmatrix} = \begin{pmatrix} 0 & 0 & 2 \\ 0 & 0 & -2\Omega \\ -\Omega & 1 & 0 \end{pmatrix}\begin{pmatrix} x^2 \\ \dot{x}^2 \\ x\dot{x} \end{pmatrix} = \boldsymbol{A}\begin{pmatrix} x^2 \\ \dot{x}^2 \\ x\dot{x} \end{pmatrix}. \tag{D.13}$$

Like in Eq. (D.6), the matrix $\boldsymbol{A}$ splits into

$$\boldsymbol{A}(t) = \boldsymbol{A}_0 + \sigma_\Omega \eta(t)\boldsymbol{A}_1 = \begin{pmatrix} 0 & 0 & 2 \\ 0 & 0 & -2\Omega_0 \\ -\Omega_0 & 1 & 0 \end{pmatrix} + \sigma_\Omega \eta(t)\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -2 \\ -1 & 0 & 0 \end{pmatrix}, \tag{D.14}$$

so that the equation for the averages becomes

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{pmatrix} \langle x^2 \rangle \\ \langle \dot{x}^2 \rangle \\ \langle x\dot{x} \rangle \end{pmatrix} = \left\{ \boldsymbol{A}_0 + \sigma_\Omega^2 \int_0^{+\infty} \langle \eta(t)\eta(t-t') \rangle \boldsymbol{B}(t')\,\mathrm{d}t' \right\}\begin{pmatrix} \langle x^2 \rangle \\ \langle \dot{x}^2 \rangle \\ \langle x\dot{x} \rangle \end{pmatrix}, \tag{D.15}$$

where $\boldsymbol{B}(t) = \boldsymbol{A}_1 \exp(\boldsymbol{A}_0 t)\boldsymbol{A}_1 \exp(-\boldsymbol{A}_0 t)$.

When the process $\eta(t)$ is Gaussian and δ-correlated, Eq. (D.15) is exact, and the integral can be computed explicitly: writing $\langle \eta(t)\eta(t-t') \rangle = \tau\delta(t')$, where τ is the correlation time scale of the random process, we obtain

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{pmatrix} \langle x^2 \rangle \\ \langle \dot{x}^2 \rangle \\ \langle x\dot{x} \rangle \end{pmatrix} = \left\{ \boldsymbol{A}_0 + \frac{\sigma_\Omega^2 \tau}{2}\boldsymbol{B}(0) \right\}\begin{pmatrix} \langle x^2 \rangle \\ \langle \dot{x}^2 \rangle \\ \langle x\dot{x} \rangle \end{pmatrix}. \tag{D.16}$$

From the definition of $\boldsymbol{B}(t)$ it follows then that $\boldsymbol{B}(0) = \boldsymbol{A}_1^2$, and by an easy calculation we find

$$\boldsymbol{A}_0 + \sigma_\Omega^2 \tau \boldsymbol{A}_1^2 = \begin{pmatrix} 0 & 0 & 2 \\ \sigma_\Omega^2 \tau & 0 & -2\Omega_0 \\ -\Omega_0 & 1 & 0 \end{pmatrix} \tag{D.17}$$

which is the result used in Section 4.1.3.

References

[1] H. Poincaré, Les méthodes nouvelles de la mécanique celeste, Gauthier-Villars, Paris, 1892.
[2] V.I. Arnol'd, Mathematical Methods of Classical Mechanics, Springer, Berlin, 1978.

[3] E. Fermi, Nuovo Cimento 25 (1923) 267. E. Fermi, Nuovo Cimento 26 (1923) 105.
[4] A.N. Kol'mogorov, Dokl. Akad. Nauk. SSSR 98 (1954) 527; V.I. Arnol'd, Russ. Math. Surv. 18 (1963) 9; J. Moser, Nachr. Akad. Wiss. Göttingen Math. Phys. Kl. 2 1 (1962) 1; see also Refs. [2,57].
[5] N.S. Krylov, Works on the Foundations of Statistical Physics, Princeton University Press, Princeton, 1979.
[6] J. Hadamard, J. Math. Pures Appl. 4 (1898) 27; G.A. Hedlund, Bull. Am. Math. Soc. 45 (1939) 241; E. Hopf, Proc. Natl. Acad. Sci. 18 (1932) 263.
[7] Ya.G. Sinai (Ed.), in: Dynamical Systems II, Encyclopædia of Mathematical Sciences, Vol. 2, Springer, Berlin, 1989.
[8] V.I. Anosov, Proc. Steklov Math. Inst. 90 (1967) 1, also reprinted in [128].
[9] Ya.G. Sinai, Sov. Math. Dokl. 4 (1963) 1818; see also Ref. [7].
[10] Y. Aizawa, J. Phys. Soc. Jpn. 33 (1972) 1693.
[11] C.P. Ong, Adv. Math. 15 (1975) 269.
[12] M.C. Gutzwiller, J. Math. Phys. 18 (1977) 806.
[13] J.F.C. van Velsen, J. Phys. A 13 (1980) 833.
[14] V.G. Gurzadyan, G.K. Savvidy, Astron. Astrophys. 160 (1986) 203.
[15] A. Knauf, Commun. Math. Phys. 110 (1987) 89.
[16] M. Szydlowski, J. Math. Phys. 35 (1994) 1850.
[17] Y. Aizawa, J. Korean Phys. Soc. 28 (1995) S310.
[18] B. Nobbe, J. Stat. Phys. 78 (1995) 1591.
[19] M. Szydlowski, M. Heller, W. Sasin, J. Math. Phys. 37 (1996) 346.
[20] H.E. Kandrup, Astrophys. J. 364 (1990) 420; Physica A 169 (1990) 73; Phys. Rev. E 56 (1997) 2722.
[21] S. Ulam, Introduction to the Los Alamos Report LA-1940, in: E. Segré (Ed.), Collected Papers of Enrico Fermi, Vol. 2, University of Chicago, Chicago, 1965, p. 978.
[22] M. Pettini, Phys. Rev. E 47 (1993) 828.
[23] L. Casetti, Laurea Thesis, Università di Firenze, 1993, unpublished.
[24] L. Casetti, M. Pettini, Phys. Rev. E 48 (1993) 4320.
[25] L. Casetti, R. Livi, M. Pettini, Phys. Rev. Lett. 74 (1995) 375.
[26] M. Cerruti-Sola, M. Pettini, Phys. Rev. E 51 (1995) 53.
[27] M. Pettini, R. Valdettaro, Chaos 5 (1995) 646.
[28] C. Clementi, Laurea Thesis, Università di Firenze, 1995, unpublished.
[29] L. Caiani, Laurea Thesis, Università di Firenze, 1995, unpublished.
[30] M. Cerruti-Sola, M. Pettini, Phys. Rev. E 53 (1996) 179.
[31] L. Casetti, C. Clementi, M. Pettini, Phys. Rev. E 54 (1996) 5969.
[32] C. Clementi, Dynamics of homopolymer chain models, MSc Thesis, SISSA/ISAS, Trieste, 1996.
[33] L. Casetti, Aspects of dynamics, geometry, and statistical mechanics in Hamiltonian systems, Ph.D. Thesis, Scuola Normale Superiore, Pisa, 1997.
[34] L. Casetti, A. Macchi, Phys. Rev. E 55 (1997) 2539.
[35] M. Cerruti-Sola, R. Franzosi, M. Pettini, Phys. Rev. E 56 (1997) 4872.
[36] L. Caiani, L. Casetti, C. Clementi, M. Pettini, Phys. Rev. Lett. 79 (1997) 4361.
[37] L. Caiani, L. Casetti, C. Clementi, G. Pettini, M. Pettini, R. Gatto, Phys. Rev. E 57 (1998) 3886.
[38] L. Caiani, L. Casetti, M. Pettini, J. Phys. A 31 (1998) 3357.
[39] L. Casetti, R. Gatto, M. Pettini, J. Phys. A 32 (1999) 3055.
[40] V.S. Dryuma, Teor. Mat. Fiz. 99 (1994) 241.
[41] P. Cipriani, M. Di Bari, Planetary Space Sci. 46 (1998) 1499; P. Cipriani, M. Di Bari, Planetary Space Sci. 46 (1998) 1543; P. Cipriani, M. Di Bari, Phys. Rev. Lett. 81 (1998) 5532; M. Di Bari, D. Boccaletti, P. Cipriani, G. Pucacco, Phys. Rev. E 55 (1997) 6448; P. Cipriani, G. Pucacco, Nuovo Cimento B 109 (1994) 325.
[42] C. Alabiso, N. Besagni, M. Casartelli, P. Marenzoni, J. Phys. A 29 (1996) 3733.
[43] R. Thom, in: S.A. Rice, K.F. Freed, J.C. Light (Eds.), Statistical Mechanics, University of Chicago Press, Chicago, 1972, p. 93.
[44] More on phase transitions and Catastrophe theory can be found, in: T. Poston, I. Stewart (Eds.), Catastrophe Theory and its Applications, Pitman Press, London, 1978, and references therein quoted.

[45] M. Rasetti, Topological concepts in the theory of phase transitions, in: H.D. Döbner (Ed.), Differential Geometric Methods in Mathematical Physics, Springer, New York, 1979; in: W. Güttinger (Ed.), Structural Stability in Statistical Mechanics, Springer Tracts in Math., Springer, New York, 1979.
[46] G. Ruppeiner, Phys. Rev. A 44 (1991) 3583.
[47] D.H.E. Gross, E. Votyakov, cond-mat/9911257 and references therein quoted.
[48] H.H. Rugh, Phys. Rev. Lett. 78 (1997) 772; J. Phys. A 31 (1998) 7761.
[49] C. Giardinà, R. Livi, J. Stat. Phys. 91 (1998) 1027.
[50] M. Cerruti-Sola, C. Clementi, M. Pettini, Phys. Rev. E 61 (2000) 5171.
[51] H. Rund, The Differential Geometry of Finsler Spaces, Springer, Berlin, 1959; G.S. Asanov, Finsler Geometry, Relativity, and Gauge Theories, Reidel, Dordrecht, 1985.
[52] K. Mohring et al., Ann. Phys. 127 (1980) 198.
[53] L.P. Eisenhart, Ann. Math. (Princeton) 30 (1929) 591.
[54] A. Lichnerowicz, Théories rélativistes de la gravitation et de l'éléctromagnétisme, Masson, Paris, 1955.
[55] J.E. Marsden, T.S. Ratiu, Introduction to Mechanics and Symmetry, Springer, New York, 1994.
[56] E. Gozzi, M. Reuter, Chaos, Solitons and Fractals 4 (1994) 1117.
[57] V.I. Arnol'd (Ed.), in: Dynamical sytsems III, Encyclopædia of Mathematical Sciences, Vol. 3, Springer, Berlin, 1988.
[58] G. Knieper, H. Weiss, J. Differential Geom. 39 (1994) 229.
[59] M.P. do Carmo, Riemannian Geometry, Birkhäuser, Boston, 1993.
[60] A.H. Nayfeh, D.T. Mook, Nonlinear Oscillations, Wiley, New York, 1979.
[61] M. Hénon, C. Heiles, Astron. J. 69 (1964) 73.
[62] M.C. Gutzwiller, Chaos in Classical and Quantum Mechanics, Springer, New York, 1990.
[63] M. Abramowitz, I.A. Stegun, Handbook of Mathematical Functions, Dover, New York, 1965.
[64] G. Gallavotti, E.G.D. Cohen, Phys. Rev. Lett. 74 (1995) 2694; J. Stat. Phys. 80 (1995) 931.
[65] S.I. Goldberg, Curvature and Homology, Dover, New York, 1965.
[66] J. Cheeger, D.G. Ebin, Comparison Theorems in Riemannian Geometry, North-Holland, Amsterdam, 1975.
[67] N.G. Van Kampen, Phys. Rep. 24 (1976) 71.
[68] E. Fermi, J. Pasta, S. Ulam (with M. Tsingou), Los Alamos Report LA-1940, reprinted in: D.C. Mattis (Ed.), The Many-Body Problem, World Scientific, Singapore, 1992, and also in: E. Segré (Ed.), Collected Papers of Enrico Fermi, Vol. 2, University of Chicago, Chicago, 1965, p. 978.
[69] R. Franzosi, R. Gatto, G. Pettini, M. Pettini, Phys. Rev. E 61 (2000) R3299.
[70] J.L. Lebowitz, J.K. Percus, L. Verlet, Phys. Rev. 153 (1967) 250.
[71] R. Livi, M. Pettini, S. Ruffo, A. Vulpiani, J. Stat. Phys. 48 (1987) 539.
[72] L. Casetti, M. Cerruti-Sola, M. Modugno, G. Pettini, M. Pettini, R. Gatto, Riv. Nuovo Cimento 22 (1999) 1.
[73] M. Pettini, M. Landolfi, Phys. Rev. A 41 (1990) 768.
[74] M. Pettini, M. Cerruti-Sola, Phys. Rev. A 44 (1991) 975.
[75] D. Escande, H. Kantz, R. Livi, S. Ruffo, J. Stat. Phys. 76 (1994) 605.
[76] L. Bunimovič, G. Casati, I. Guarneri, Phys. Rev. Lett. 77 (1996) 2941.
[77] E.G.D. Cohen, Am. J. Phys. 58 (1990) 619.
[78] Rigorous results about the appearance of these singularities have been given in: C.N. Yang, T.D. Lee, Phys. Rev. 87 (1952) 404, as far as the grand-canonical ensemble is concerned, and in: D. Ruelle, Thermodynamic Formalism, Encyclopaedia of Mathematics and its Applications, Addison-Wesley, New York, 1978, as far as the canonical ensemble is concerned.
[79] M.E. Fisher, D. Jasnow, in: C. Domb, M.S. Green (Eds.), Phase Transitions and Critical Phenomena, Vol. 4, Academic Press, London, 1976; M.E. Fisher, Rev. Mod. Phys. 70 (1998) 653.
[80] K.G. Wilson, J. Kogut, Phys. Rep. 12 (1974) 75.
[81] S.-K. Ma, Modern Theory of Critical Phenomena, Benjamin, New York, 1976.
[82] G. Parisi, Statistical Field Theory, Addison-Wesley, New York, 1988.
[83] J. Zinn-Justin, Quantum Field Theory and Critical Phenomena, Oxford University press, Oxford, 1989.
[84] M. Le Bellac, Quantum and Statistical Field Theory, Oxford University Press, Oxford, 1991.

[85] N. Goldenfeld, Lectures on Phase Transitions and the Renormalization Group, Addison-Wesley, New York, 1992.
[86] P. Butera, G. Caravati, Phys. Rev. A 36 (1987) 962.
[87] A. Bonasera, V. Latora, A. Rapisarda, Phys. Rev. Lett. 75 (1995) 3434.
[88] C.S. O'Hern, D.A. Egolf, H.S. Greenside, Phys. Rev. E 53 (1996) 3374.
[89] Ch. Dellago, H.A. Posch, W.G. Hoover, Phys. Rev. E 53 (1996) 1485.
[90] Ch. Dellago, H.A. Posch, Physica A 230 (1996) 364.
[91] Ch. Dellago, H.A. Posch, Physica A 237 (1997) 95.
[92] Ch. Dellago, H.A. Posch, Physica A 240 (1997) 68.
[93] V. Mehra, R. Ramaswamy, Phys. Rev. E 56 (1997) 2508.
[94] V. Latora, A. Rapisarda, S. Ruffo, Phys. Rev. Lett. 80 (1998) 692.
[95] M.-C. Firpo, Phys. Rev. E 57 (1998) 6599.
[96] V. Latora, A. Rapisarda, S. Ruffo, Physica D 131 (1999) 38.
[97] M. Antoni, S. Ruffo, A. Torcini, Dynamics and Statistics of simple models with infinite-range attractive interactions, cond-mat/9908336.
[98] R. Livi, A. Maritan, Phys. Rev. D 23 (1981) 2252.
[99] M. Antoni, S. Ruffo, Phys. Rev. E 52 (1995) 2361.
[100] Y.Y. Yamaguchi, Prog. Theor. Phys. 95 (1996) 717.
[101] C. Anteneodo, C. Tsallis, Phys. Rev. Lett. 80 (1998) 5313.
[102] R. Franzosi, L. Casetti, L. Spinelli, M. Pettini, Phys. Rev. E 60 (1999) R5009.
[103] R. Franzosi, Aspetti geometrici e topologici nello studio delle transizioni di fase, Ph.D. Thesis, Università di Firenze, 1998.
[104] L. Casetti, E.G.D. Cohen, M. Pettini, Phys. Rev. Lett. 82 (1999) 4160.
[105] M. Spivak, A Comprehensive Introduction to Differential Geometry, Publish or Perish, Berkeley, 1979.
[106] J.A. Thorpe, Elementary Topics in Differential Geometry, Springer, New York, 1979.
[107] R. Franzosi, M. Pettini, L. Spinelli, Phys. Rev. Lett. 84 (2000) 2774.
[108] This classic co-area formula can be found in: H. Federer, Geometric Measure Theory, Springer, Berlin, 1969.
[109] M. Nakahara, Geometry, Topology and Physics, Adam Hilgher, Bristol, 1991.
[110] K. Binder, MonteCarlo Methods in Statistical Physics, Springer, Berlin, 1979.
[111] To solve Eqs. (135) for a general (i.e., not mean-field-like) potential energy function one could borrow methods from: J. Vollmer, W. Breymann, R. Schilling, Phys. Rev. B 47 (1993) 11767.
[112] C. Nash, Differential Topology and Quantum Field Theory, Academic, London, 1991.
[113] E.G.D. Cohen, L. Rondoni, Chaos 8 (1998) 357.
[114] A. Dhar, Physica A 259 (1989) 119.
[115] L. Spinelli, Une approche topologique des transitions de phase, Ph.D. Thesis, Université de Provence, 1999.
[116] L. Spinelli, R. Franzosi, M. Pettini, in preparation.
[117] D.H.E. Gross, Phys. Rep. 279 (1997) 119.
[118] L.D. Landau, E.M. Lifšits, The Classical Theory of Fields, Pergamon, London, 1968.
[119] R.M. Wald, General Relativity, The University of Chicago Press, Chicago, 1984.
[120] S. Kobayashi, K. Nomizu, Foundations of Differential Geometry, Wiley, New York, 1991.
[121] M. Morse, Calculus of Variations in the Large, American Mathematical Society, Providence, RI, 1934.
[122] J. Milnor, Morse Theory, Annals of Mathematical Studies, Vol. 51, Princeton University Press, Princeton, 1963.
[123] R.S. Palais, C. Terng, Critical Point Theory and Submanifold Geometry, Lecture Notes on Mathematics, Vol. 1353, Springer, Berlin, 1988.
[124] R.P. Feynman, R.B. Leighton, M. Sands, The Feynman Lectures on Physics, Vol. III, Addison-Wesley, Reading, 1965.
[125] A.J. Lichtenberg, M.A. Lieberman, Regular and Chaotic Dynamics, 2nd edition, Springer, New York, 1992.
[126] M. Tabor, Chaos and Integrability in Nonlinear Dynamics, Wiley, New York, 1989.
[127] E. Ott, Chaos in Dynamical Systems, Cambridge University Press, Cambridge, 1993.
[128] R.S. MacKay, J.D. Meiss (Eds.), Hamiltonian Systems: a Reprint Selection, Adam Hilger, Bristol, 1990.
[129] V.K. Mel'nikov, Trans. Moscow Math. Soc. 12 (1963) 1.

[130] V.I. Arnol'd, Russ. Math. Surv. 18 (1964) 85, also reprinted in [128].
[131] J.-P. Eckmann, D. Ruelle, Rev. Mod. Phys. 57 (1985) 617.
[132] V.I. Oseledeč, Trans. Moscow Math. Soc. 19 (1969) 197.
[133] G. Benettin, L. Galgani, J.-M. Strelcyn, Phys. Rev. A 14 (1976) 2338; G. Benettin, L. Galgani, A. Giorgilli, J.-M. Strelcyn, Meccanica 15 (1980) 1.
[134] S. Ruffo, Lyapunov spectra and characterization of chaotic dynamics, in: R. Livi, J.-P. Nadal (Eds.), Complex Dynamics, Nova Publishing, 1994.
[135] C.P. Dettmann, G.P. Morriss, Phys. Rev. E 55 (1997) 3693 and references therein.
[136] R. Livi, A. Politi, S. Ruffo, J. Stat. Phys. 19 (1986) 2083; Ch. Dellago, L. Glotz, H.A. Posch, Phys. Rev. E 52 (1995) 4817; see also Ref. [92].
[137] Ya.G. Sinai, J. Bifur. Chaos Appl. Sci. Eng. 6 (1996) 1137.